

**L'IMPACT DU DIALOGUE HOMME-MACHINE CHEZ
LES PERSONNES ÂGÉES DANS LE CADRE D'UNE
AIDE À DOMICILE**

par

Léopold Lieb

Mémoire présenté au Département d'informatique
en vue de l'obtention du grade de maîtrise en informatique (M.Sc.)

FACULTÉ DES SCIENCES

UNIVERSITÉ DE SHERBROOKE

Sherbrooke, Québec, Canada, 14 juin 2018

Le 14 juin 2018,

le jury a accepté le mémoire de Monsieur Léopold Lieb dans sa version finale.

Membres du jury:

Professeur Hélène Pigot
Directrice de recherche
Département de science

Professeur Frodual Kabanza
Président rapporteur
Département de science

Professeur Bessam Abdulrazak
Membre interne
Département de science

Sommaire

Ce mémoire porte sur la conception et l'évaluation de solutions technologiques d'assistance aux personnes vieillissantes par un aspect vocal. Le nombre des personnes âgées va continuer d'augmenter durant les prochaines années et de plus en plus de nouvelles technologies se développent pour l'assistance aux aînés. Ces technologies se développent car les aînés souhaitent rester indépendants et vivre dans leur milieu préféré [1]. De plus, ces technologies offrent un accès à une assistance à un coût plus viable, pour les aînés ou leur famille, que par une assistance à travers une aide médicale ou un placement dans des centres spécialisés [2] [3]. Cependant, utiliser des technologies nécessite de les adapter pour toujours rester en accord avec les besoins des aînés [4], [5]. Des études ont déjà montré l'apport d'une interaction vocal avec un agent virtuel pour les aînés, que ça soit pour les guider et les assister, ou donner un aspect plus humain, réduisant la distance entre la technologie et les personnes âgées [6]. L'utilisation d'une interface tactile peut apporter des difficultés telle qu'une mauvaise compréhension de composants ou de la difficulté pour sélectionner des options [7], c'est pourquoi le vocal permet à l'utilisateur de formuler une demande sans se soucier de comprendre le fonctionnement tactile qui peut y être lié. L'objectif de cette recherche est de fournir un dialogue entre une personne et un agent virtuel dans le cadre de la commande d'un calendrier. L'utilisateur peut contrôler diverses fonctionnalités d'un calendrier à travers différentes formulations. Dans l'étude de Ramin Yaghoubzadeh et al. [6], la discussion avec l'agent virtuel était centré simplement sur l'ajout de rendez-vous. Cependant, cette étude montre que les aînés ont besoin de retours vocaux fréquents pour leur permettre de suivre l'activité en cours. De plus, l'utilisateur est libre d'engager une conversation avec l'agent virtuel sur des sujets divers. Des solutions comme Siri, ou Ok Google n'instaurent pas un dialogue avec

SOMMAIRE

l'utilisateur dès lors que ce dernier fait une demande plus générale. L'agent personnel se contente de faire une recherche internet et de donner à l'utilisateur des liens en rapport avec sa demande [8]. Le travail de cette recherche se base sur une approche de conception centrée utilisateur, participative et pluridisciplinaire [9], faisant appel à des connaissances sur l'ergonomie et le dialogue pour faciliter l'interaction Homme-Machine, et l'informatique pour développer des technologies d'assistance. L'interaction Homme-Machine s'appuie sur deux modes de dialogues :

- Un dialogue par règles permettant de contrôler le calendrier par des commandes vocales. Il existe déjà des modèles par règles comme CleverBot [10], mais peu sont centré sur l'assistance aux aînés. Le dialogue par règles a ici à la fois pour but de fournir un contrôle d'une application comme certains modèles déjà existants, mais d'apporter en plus une aide adaptée aux aînés. L'âge joue sur plusieurs aspects cognitifs, comme la vitesse de compréhension, l'agilité mentale et la mémoire [11]. C'est pourquoi le dialogue doit être plus fourni pour permettre aux aînés d'avoir plus de retours et de savoir à chaque fois où en est rendu l'agent virtuel avec lequel ils dialoguent, pour à la fois être sûr d'être compris et à la fois de comprendre ce qui est en cours. ;
- Un dialogue libre utilisant des réseaux de neurones pour offrir la possibilité à l'utilisateur de discuter de sujets libres. Les applications, comme celle dans l'étude de Ramin Yaghoubzadeh et al. [6], se focalisent principalement sur l'assistance aux aînés à travers le dialogue, et se servent d'un agent virtuel pour apporter un aspect plus humain. Cependant, une interaction naturelle passe aussi par des échanges libres qui n'ont pas de rapport avec des demandes spécifiques. L'importance du dialogue libre est de réduire la distance homme-machine à laquelle font face certains aînés [12].

En plus de l'utilisation de dialogues, tout en restant dans l'idée d'assister et de permettre aux aînés de toujours savoir où ils en sont, l'utilisation d'une interface graphique est nécessaire [13]. Lier une interface tactile à une modalité vocale permet aux aînés d'avoir à la fois un retour vocal ainsi qu'un retour visuel. Le retour visuel est d'autant plus important pour ces personnes puisqu'elles peuvent faire face à des problèmes de mémoire [11]. L'interaction vocale a été testée selon deux expérimentations :

- Une première expérimentation à domicile pour valider le dialogue par règles ;
- Une deuxième expérimentation par rencontres avec les participants et à domicile pour valider l'évolution du dialogue par règles, l'apport d'un dialogue libre et la communication entre la modalité vocale et la modalité tactile ;

Les résultats ont montré que les aînés interagissent principalement avec le calendrier en le tactile. Le vocal est moins utilisé par un manque de performance du dialogue. Néanmoins, la modalité vocale couplée à la modalité tactile permet à l'utilisateur d'adapter ses actions. Le tactile est privilégié pour des actions courtes comme la navigation, et le vocal est privilégié pour des actions plus complexes. Les aînés se disent prêts à utiliser le vocal s'il est plus performant. L'agent virtuel doit aider les aînés dans leurs tâches, leur expliquer des actions et savoir discuter avec eux.

Mots-clés: Vieillissement ; autonomie ; agent virtuel ; dialogue homme-machine ; dialogue libre

Remerciements

Je souhaite remercier Hélène Pigot, co-fondatrice du laboratoire DOMUS et directrice de la présente recherche, de m’avoir offert l’opportunité de faire ce travail. Je désire aussi la remercier pour sa confiance, son soutien et ses conseils.

De plus, je remercie l’équipe du laboratoire DOMUS pour leur support tout au long de ma maîtrise, et plus spécialement Amadine Porcher et Wathek Bellah Loued avec qui j’ai partagé ce projet.

Finalement, je souhaite remercier mes parents, Sophie André et Jean-Marie Lieb, ainsi que ma sœur Marie-Anne Lieb, qui m’ont soutenu et encouragé tout au long de mon mémoire.

REMERCIEMENTS

Abréviations

AVQ Activité de la Vie Quotidienne

CNN Réseau neuronal convolutif ou Convolutional Neural Network

DHM Dialogue Homme-Machine

DNN Réseau de neurones d'apprentissage profond ou Deep Neural Network

E2E End-To-End

GUI Interface Graphique ou Graphic User Interface

IUV Interaction Utilisateur Vocale

LDA Latent Dirichlet allocation

MII Méthode incrémentale et itérative

MoCA Montreal Cognitive Assessment

O1 Objectif 1

RNN Réseaux de neurones récurrents ou recurrent neural network

RVI Réponse Vocale Interactive

seq2seq sequence to sequence

SO Sous-objectif

TALN Traitement automatique du langage naturel

TUI Environnement textuel ou Text-based user interface

UML Langage de Modélisation Unifié

Table des matières

Sommaire	ii
Remerciements	v
Abréviations	vii
Table des matières	viii
Liste des figures	xii
Liste des tableaux	xv
Introduction	1
1 Revue de littérature	4
1.1 Personnes âgées et technologies	4
1.2 Importance du dialogue homme-machine	5
1.2.1 Force et limites des interactions utilisateur vocal (IUV)	5
1.2.2 Typologie des dialogues	6
1.2.3 Composants d'un dialogue vocal	7
1.2.4 Étapes du design du dialogue	9
1.2.5 Avoir un dialogue structuré	13
1.2.6 Retour du système sur la demande de l'utilisateur	14
1.3 Différentes solutions pour le DHM	19
1.3.1 Modèles à base de reconnaissance de motifs	20
1.3.2 Modèles pour un dialogue général	21

TABLE DES MATIÈRES

2	Objectifs et méthodologie	24
2.1	Contexte du calendrier Amelis	24
2.1.1	Buts généraux du calendrier Amelis	24
2.1.2	Étapes de conception du calendrier Amelis	25
2.2	Fonctionnalités et aspect du calendrier Amelis	26
2.3	Objectifs de cette recherche	29
2.4	Méthodologie	31
2.4.1	Utiliser des règles pour contrôler l'application Amelis et générer des réponses pour l'agent virtuel	31
2.4.2	Utiliser un réseau de neurones pour un dialogue libre	34
2.4.3	Synchroniser le vocal et le tactile	38
2.4.4	Utiliser des tests pour vérifier sur un aspect technique le code et valider le fonctionnement du calendrier avec des utilisateurs	39
3	Conception	44
3.1	Conception d'une architecture adaptable pour le dialogue et l'agent virtuel au sein de l'architecture d'Amelis	44
3.1.1	Architecture tactile d'Amelis	44
3.1.2	Architecture vocale d'Amelis	45
3.2	Conception du dialogue par règles	46
3.2.1	Écriture des règles selon les fonctionnalités vocales de l'application Amelis	46
3.2.2	Mise en place des synonymes	48
3.2.3	Traitement des dates et des heures par le dialogue par règles	49
3.2.4	Augmentation des choix d'actions	52
3.2.5	Impasse du dialogue par règles sur les sujets libres	59
3.3	Création d'un dialogue libre à l'aide d'un réseau de neurones	60
3.3.1	Utilisation de deux réseaux de neurones libres d'accès	60
3.3.2	Choix d'un corpus	61
3.3.3	Variation des paramètres	63
3.4	Conception de la liaison entre le vocal et le tactile	64
3.4.1	Synchronisation du vocal vers le tactile	66

TABLE DES MATIÈRES

3.4.2	Synchronisation du tactile vers le vocal	67
3.5	Conception des tests de validation des fonctions de l'application Amelis	68
3.5.1	Utilisation de tests end-to-end (E2E) pour le dialogue par règles	69
3.5.2	Évaluation des apprentissages des réseaux de neurones	70
4	Résultats	73
4.1	Résultats aux tests techniques	73
4.1.1	Les tests E2E	73
4.1.2	Le test BLEU et le test sémantique	77
4.2	Résultats de la première expérimentation	77
4.2.1	Présentation des participants	78
4.2.2	Taux de réussite des actions vocales des participants	81
4.2.3	Causes des erreurs vocales	83
4.2.4	Utilisation tactile et vocale durant les 3 mois d'expérimentation	87
4.3	Résultats de la deuxième expérimentation	89
4.3.1	Présentation des participants	94
4.3.2	Résultats des rencontres entre les participants et l'expérimentateur	94
4.3.3	Taux de réussite des actions vocales	97
4.3.4	Comparaison de l'utilisation du tactile et du vocal par les participants	100
4.3.5	Utilisation du dialogue libre	104
4.3.6	Évaluation de l'application Amelis par les participants	106
5	Discussion	108
5.1	Objectif 2 : Offrir un contrôle de l'application Amelis par la parole . .	108
5.1.1	Contrôle de l'application selon un dialogue par règles	108
5.1.2	Formulations diverses des commandes dans le dialogue par règles	109
5.1.3	Demandes intelligibles de l'agent virtuel	110
5.2	Objectif 3 : Offrir des contextes de discussion libre entre l'agent virtuel et l'utilisateur pour donner un aspect plus humain lors des échanges .	111
5.2.1	Réseaux de neurones et dialogue libre	111

TABLE DES MATIÈRES

5.2.2	Qualité du dialogue libre et acceptabilité du dialogue de l'agent virtuel	112
5.3	Objectif 4 : Intégrer un mode d'interaction tactile et un mode d'interaction vocal pour offrir un choix d'interaction plus large et plus simple à l'utilisateur	114
5.3.1	Deux modes d'interaction élargissent les possibilités de choix d'interaction	114
5.3.2	Synchronisation entre les deux modes d'interactions	115
5.4	Objectif 5 : Évaluer les performances techniques et l'impact sur l'utilisateur du dialogue de contrôle et libre	116
5.4.1	La reconnaissance vocale comprend l'utilisateur	116
5.4.2	Validation par les tests techniques	117
5.4.3	Appréciation du dialogue avec l'agent virtuel par les utilisateurs	118
Conclusion		120
A Maquettes du calendrier Amelis		123
Bibliographie		127

Liste des figures

1.1	Exemple de la structure du dialogue autour de la prise de rendez-vous	14
1.2	Exemple de retours visuels lors de l'ajout d'un contact	17
1.3	Cohabitation de deux types de dialogue	19
1.4	Système de règles et de priorités (ici dièse plus important que l'accent circonflexe)	21
2.1	Maquette commune du calendrier Amelis entre le Québec et la France	27
2.2	Étapes pour créer une structure pour les règles pour récupérer les in- formations de l'utilisateur	32
2.3	Étapes pour créer une structure pour les règles pour fournir des ré- ponses à l'utilisateur	33
2.4	Étapes pour adopter le dialogue à utiliser pour les règles	34
2.5	Étapes pour diversifier les interactions vocales pour simplifier les choix d'actions	35
2.6	Étapes pour déterminer le réseau de neurones à utiliser	35
2.7	Étapes pour déterminer le corpus libre à utiliser	36
2.8	Étapes pour mettre en place la traduction et déterminer si c'est une solution viable	37
2.9	Étapes pour la synchronisation du vocal au tactile	38
2.10	Étapes pour la synchronisation du tactile au vocal	39
2.11	Étapes pour tester le dialogue par règles et la synchronisation avec des tests E2E	40
3.1	Architecture du tactile	45

LISTE DES FIGURES

3.2	Architecture du vocal	46
3.3	Architecture du dialogue par règles avec le traitement des synonymes	49
3.4	Étapes pour traiter les dates avec le dialogue par règles	53
3.5	Architecture du dialogue par règles avec le traitement des dates . . .	54
3.6	Structure des règles avec un traitement au cas par cas	56
3.7	Structure des règles avec une imbrication des règles	57
3.8	Architecture d'un encodeur décodeur	63
3.9	Convergence de la fonction d'apprentissage	64
3.10	Architecture du dialogue libre	65
3.11	Exemple de synchronisation du vocal vers le tactile	66
3.12	Architecture du vocal comprenant la synchronisation du vocal vers le tactile	67
3.13	Architecture du tactile comprenant la synchronisation du tactile vers le vocal	68
3.14	Exemple de vérification du dialogue par règles et de la synchronisation	71
4.1	Exemple d'erreur suite à un test	74
4.2	Exemple du déroulement d'un test E2E d'ajout sous la forme d'un diagramme UML	75
4.3	Exemple de tests sur un ajout simple de rendez-vous	76
4.4	Présentation de la participante P1_E1	78
4.5	Présentation de la participante P2_E1	79
4.6	Présentation de la participante P3_E1	80
4.7	Présentation du participant P4_E1	80
4.8	Nombre d'actions vocales réussies de la participante P1_E1	81
4.9	Nombre d'actions vocales réussies de la participante P2_E1	82
4.10	Nombre d'actions vocales réussies de la participante P3_E1	82
4.11	Nombre d'actions vocales réussies du participant P4_E1	83
4.12	Causes des erreurs vocales de la participante P1_E1	84
4.13	Causes des erreurs vocales de la participante P2_E1	85
4.14	Causes des erreurs vocales de la participante P3_E1	86
4.15	Causes des erreurs vocales du participant P4_E1	86

LISTE DES FIGURES

4.16	Utilisation du tactile et du vocal selon les fonctionnalités de l'application Amelis de la participante P1_E1	90
4.17	Utilisation du tactile et du vocal selon les fonctionnalités de l'application Amelis de la participante P2_E1	91
4.18	Utilisation du tactile et du vocal selon les fonctionnalités de l'application Amelis de la participante P3_E1	92
4.19	Utilisation du tactile et du vocal selon les fonctionnalités de l'application Amelis du participant P4_E1	93
4.20	Évaluation des quatre tâches d'ajout par l'utilisatrice P1_E2	95
4.21	Évaluation des quatre tâches d'ajout par l'utilisatrice P2_E2	96
4.22	Évaluation des quatre tâches d'ajout par l'utilisatrice P3_E2	96
4.23	Évaluation des quatre tâches d'ajout par l'utilisateur P4_E2	97
4.24	Nombre d'actions vocales réussies par la participante P1_E2	98
4.25	Nombre d'actions vocales réussies par la participante P2_E2	98
4.26	Nombre d'actions vocales réussies par la participante P3_E2	99
4.27	Nombre d'actions vocales réussies par le participant P4_E2	99
4.28	Utilisation tactile et vocale de l'application Amelis par la participante P1_E2	100
4.29	Utilisation tactile et vocale de l'application Amelis par la participante P2_E2	101
4.30	Utilisation tactile et vocale de l'application Amelis par la participante P3_E2	102
4.31	Utilisation tactile et vocale de l'application Amelis par la participante P4_E2	102
4.32	Synchronisations observées sur l'ajout de rendez-vous	103
A.1	Maquette du calendrier Amelis en France	124
A.2	Maquette du calendrier Amelis au Québec	125
A.3	Maquette commune du calendrier Amelis	126

Liste des tableaux

1.1	Exemple d'un dialogue maître-maître et d'un dialogue maître-esclave	11
1.2	Exemple de discussions avec et sans changement de sujet	12
1.3	Comparaison d'une confirmation implicite et d'une confirmation générale	16
4.1	Fréquence d'utilisation tactile et vocale au cours des 3 mois d'expérimentation de la participante P1_E1	87
4.2	Fréquence d'utilisation tactile et vocale au cours des 3 mois d'expérimentation de la participante P2_E1	87
4.3	Fréquence d'utilisation tactile et vocale au cours des 3 mois d'expérimentation de la participante P3_E1	88
4.4	Fréquence d'utilisation tactile et vocale au cours des 3 mois d'expérimentation du participant P4_E1	88
4.5	Évaluation de l'application Amelis par les participants	107

LISTE DES TABLEAUX

Introduction

Au cours de ces dernières années, les nouvelles technologies, telles qu’Internet, les téléphones portables ou encore les ordinateurs ont grandement évolué et sont devenus des outils incontournables dans notre vie quotidienne [14]. Mais le fait de leur perpétuelle évolution implique un temps d’adaptation et un effort cognitif plus important [15]. Un autre phénomène marquant, de ces dernières années, est l’augmentation de la population vieillissante [16].

Des technologies pour l’aide à domicile se répandent de plus en plus [17]. Il existe également des aides pour comprendre ces technologies et pour pallier les déficiences physiques ou cognitives que développent les utilisateurs. Les outils intelligents, tant au niveau physique que cognitif sont une réelle nécessité pour permettre aux personnes âgées de retrouver une autonomie dans leur vie quotidienne. Ces appareils connectés ont pour but d’aider l’utilisateur à accéder plus facilement et plus rapidement à l’information recherchée. Le premier obstacle, auquel sont confrontées les personnes âgées, est l’interaction avec l’appareil. L’ergonomie est un point important pour permettre une meilleure compréhension des nouvelles technologies pour des personnes âgées ou avec des déficits. Selon Beaudouin-Lafon [18], le paradoxe des interfaces homme-machine est que les meilleures interfaces sont celles qui sont invisibles, celles que l’utilisateur ne remarque pas et avec lesquelles il trouve l’interaction naturelle. La communication homme-machine moderne se développe selon deux axes majeurs (1) l’interaction tactile [19] et (2) l’interaction vocale [20].

Notre travail étant centré sur des personnes vieillissantes, cela oriente vers un design simplifié [21]. Notre projet, étant principalement construit au tour d’un calendrier pour aider à la prise de rendez-vous, il ne peut faire appel à un design classique de calendrier, tel que google calendar [22]. Bien que ce dernier puisse paraître simple

d'utilisation pour des personnes familières avec la technologie, il inclut beaucoup des fonctionnalités, dont certaines peu intuitives, le rendant trop complexe pour le type de personnes que nous visons.

Une autre approche de notre projet est la mise en place d'un dialogue entre l'utilisateur et un agent virtuel. Les personnes âgées font souvent face à des difficultés concernant l'utilisation des nouvelles technologies [4]. Des solutions ont donc été mises en place, telles que des ateliers avec des aidants [23], ou l'aide venant de membres ou d'amis de la famille. L'agent virtuel a donc pour but de guider l'utilisateur, de l'aider dans ses tâches et de simplifier l'utilisation du calendrier, tout cela en permettant à l'utilisateur de formuler ses requêtes oralement, plutôt que d'utiliser le tactile, permettant ainsi de réduire la charge cognitive de l'utilisateur [15].

Thématique de recherche

Notre recherche a pour but de permettre aux personnes âgées de conserver ou de développer leur autonomie en mettant à leur disposition une technologie intelligente. La proposition d'un calendrier interactif intégrant un agent virtuel, servant d'aidant, apporterait des services d'assistance adaptés au contexte d'utilisation.

Nous nous centrons ici sur une utilisation du calendrier au sein du domicile.

Amelis est un calendrier électronique mural développé au sein du laboratoire de recherche DOMUS de l'Université de Sherbrooke [24]. Le calendrier Amelis consiste en une application installée sur une tablette tactile qui intègre un agent virtuel.

Amelis offre différentes fonctionnalités tant au niveau du traitement des rendez-vous que de la navigation au sein du calendrier. Les fonctionnalités qui y sont liées sont des actions précises qui demandent d'avoir une structure bien définie si l'on souhaite interagir en vocal. Pour cela, la mise en place d'un dialogue par règles [25] est nécessaire. Le dialogue par règles permet de différencier les actions et de traiter les informations en conséquence. Ce processus passe par un traitement de la phrase de l'utilisateur pour la décomposer selon les informations qu'il faut retenir, on parle alors de pattern matching [26]. Un deuxième type de dialogue doit être mis en place, il s'agit d'un dialogue libre. En effet, l'utilisateur peut avoir pour but de converser avec l'agent virtuel sans vouloir faire de requête. Ce dialogue permet à l'utilisateur de discuter librement avec l'agent virtuel sans pour autant chercher à effectuer une tâche spécifique sur le calendrier.

INTRODUCTION

Objectifs de recherche

Le projet Amelis vise à permettre aux aînés de préserver une autonomie dans leur vie quotidienne en adaptant notre solution pour permettre aux aînés d'avoir des interactions simples et compréhensibles avec le calendrier. Dans cet objectif global s'intègre l'objectif de cette recherche qui est de fournir un dialogue entre une personne et un agent virtuel dans le cadre de la commande d'un calendrier.

Organisation du document

Ce mémoire est organisé de la manière suivante :

1. La revue de littérature sur le dialogue homme-machine ;
2. Les objectifs et la méthodologie de cette recherche ;
3. La conception du dialogue homme-machine et la vérification de ce dialogue ;
4. Les résultats techniques sur l'application et l'évaluation de l'application par les utilisateurs ;
5. La discussion des résultats ;
6. La conclusion sur l'aboutissement de cette recherche.

Chapitre 1

Revue de littérature

1.1 Personnes âgées et technologies

L'évolution démographique se dirige vers une population plus âgée. Le taux de fertilité en Europe et celui au Canada sont tels qu'une structure de population de plus en plus âgée se met en place. Il devient nécessaire de réfléchir à de nouveaux moyens de prise en charge des aînés. Des technologies pour l'aide à domicile se répandent de plus en plus [17]. Les aînés souhaitent être le plus possible indépendants et rester à domicile. Les nouvelles technologies le permettent grâce à un coût économique [2] viable. Toutefois, il faut s'assurer d'une acceptabilité sociale surtout par les personnes âgées. En effet, le caractère impersonnel, la complexité ou la vitesse d'évolution de ces dispositifs contrastent avec l'exigence de qualité des services qui doivent être rendus à cette population aux besoins très spécifiques. L'écart générationnel avec la technologie oblige à adapter les technologies et leurs interactions [4], [5]. Beaucoup de travaux sont réalisés par rapport au design de ces technologies pour qu'elles soient plus facilement acceptées par les aînés. Une des solutions proposées pour une aide à domicile pour les personnes âgées est la communication avec un agent virtuel [27] [6]. Par exemple, l'agent virtuel Alexa [8] allume la télé, fait des réservations, fait des rappels et raconte des blagues. Cependant, utiliser une telle solution requiert qu'elle soit acceptée par les personnes devant s'en servir [28] et que ces personnes puissent interagir de manière

1.2. IMPORTANCE DU DIALOGUE HOMME-MACHINE

simple avec l'agent virtuel, tout en obtenant un résultat satisfaisant à leurs attentes.

1.2 Importance du dialogue homme-machine

Une question importante est de savoir comment une personne âgée et/ou non experte avec la technologie peut interagir de manière naturelle et intuitive. L'interaction vocale est le moyen le plus commun d'interagir avec quelqu'un ou une machine et d'exprimer des besoins [29]. Des problèmes, pouvant être évités, rencontrés lors de l'utilisation d'une interface tactile, telle qu'une mauvaise compréhension de composants ou la difficulté pour sélectionner des options [7]. Cependant, les aînés font souvent face à des problèmes de mémoire, des problèmes cognitifs et/ou des problèmes de compréhension. Il est donc important d'identifier les besoins des personnes âgées. Dans l'étude de Godde et al. [30], il est démontré que les aînés ont plus de difficultés à se faire comprendre lors d'une interaction vocale, car ils peuvent faire face à des problèmes tels que la fragmentation de phrases, un faux début de phrase ou une interruption dans l'élocution de phrases. L'âge joue sur plusieurs aspects cognitifs, comme la vitesse de compréhension, l'agilité mentale et la mémoire [11]. Néanmoins, les interfaces vocales peuvent réduire certaines difficultés quotidiennes auxquelles font face les aînés, et sont considérées comme des solutions moins intrusives [31].

1.2.1 Force et limites des interactions utilisateur vocal (IUV)

L'interaction utilisateur vocal (IUV) offre plusieurs avantages :

- La vitesse. Il est plus rapide d'énoncer une idée que de l'écrire comme l'explique Adams K. [32]. Cela s'applique encore plus pour les aînés qui peuvent être victimes de tremblements, d'un retard envers la technologie ou encore d'un mauvais contact de leur peau avec l'écran [33];
- Avoir les mains libres. Cela peut permettre de se déplacer tout en parlant à l'IUV. Par exemple, une application vocale destinée à aider à suivre des recettes pendant qu'une personne cuisine [34];
- L'intuition. Il est parfois plus simple de parler que d'avoir à interagir en tactile avec une technologie qui ne nous est pas familière ;

- L’empathie. Le vocal permet d’intégrer un ton, une sonorité sur un message. Un message écrit est souvent interprété d’une autre manière que celle souhaitée. Cependant, il existe aussi des inconvénients à l’utilisation d’IUV :
- La vie privée. Est-ce qu’une conversation peut être enregistrée ? Dans quelles mesures l’utilisateur souhaite garder ses informations privées ? Il faut toujours prendre en compte le choix de l’utilisateur sur ce qu’il veut partager [35];
- L’inconfort avec la technologie. Il n’est pas toujours aisé de parler avec une machine, qui plus est pour des aînés qui ne sont pas habitués aux nouvelles technologies [36];
- La reconnaissance vocale. La reconnaissance de la parole peut être troublée par plusieurs facteurs tels que le bruit environnant, la prononciation de l’utilisateur ou encore le temps d’écoute (quand commencer à écouter et quand arrêter).

Avant de rentrer plus en profondeur dans l’utilisation d’IUV, il est nécessaire de choisir la relation que l’utilisateur aura avec la machine.

1.2.2 Typologie des dialogues

Daniel Luzzati propose une typologie des dialogues homme-machine (DHM) [37]. La liste suivante reprend cette typologie :

- Un dialogue où l’utilisateur est maître et le système est esclave. Les systèmes à réponse vocale interactive (RVI) [38] en sont un exemple. Un répondeur automatique de banque qui demande à l’utilisateur de taper un chiffre pour faire un choix et un dialogue où le système est esclave et l’utilisateur maître, car l’utilisateur veut acquérir un certain nombre d’informations qu’il sait être connues du système ;
- Un dialogue où l’utilisateur est maître et le système est serviteur. Dans ce cas, l’utilisateur ne sait pas exactement ce qu’il veut ou comment l’obtenir. Le système est serviteur, car il a pour but d’essayer de faire des propositions à l’utilisateur pour trouver ce qu’il cherche. Par exemple, l’utilisateur cherche une rue, le système essaie de trouver la ville, un quartier, un magasin... en posant des questions à l’utilisateur pour trouver le nom de la rue ;
- Un dialogue où l’utilisateur et le système sont maîtres. Cela a lieu dans le cas

1.2. IMPORTANCE DU DIALOGUE HOMME-MACHINE

d'un dialogue naturel ou d'un dialogue libre. L'utilisateur et le système parlent d'égaux à égaux. Leur but est simplement d'instaurer un échange.

Plusieurs types de dialogues peuvent donc être mis en place entre l'utilisateur et la machine selon les besoins. Mais il est nécessaire de se demander ce qui englobe la mise en place d'un dialogue.

1.2.3 Composants d'un dialogue vocal

Pour concevoir un dialogue, il faut répondre à ces trois caractéristiques :

- La reconnaissance vocale ;
- La correction de la sortie ;
- Les tours de parole.

1.2.3.1 Reconnaissance vocale

La reconnaissance vocale s'est grandement améliorée ces dernières années [39], tant au niveau des mots à reconnaître qu'au niveau du traitement des nuisances sonores externes [40]. La reconnaissance vocale peut même associer des émotions selon l'intonation de la voix [41]. Les systèmes de reconnaissance vocale tentent d'atteindre ou de dépasser le niveau de reconnaissance vocale d'un humain [42]. Il existe de nombreux modules pour réaliser de la reconnaissance vocale, ce qui oblige à faire une sélection selon nos besoins :

- Utiliser un logiciel libre d'utilisation et gratuit ;
- Avoir un module de reconnaissance suffisamment entraîné, comme Google Speech API, Siri ou Cortana [43];
- Avoir un module capable de comprendre le français ;
- Pouvoir améliorer le logiciel de reconnaissance pour traiter des exceptions et l'adapter aux utilisateurs.

Tout système demande une adaptation à nos besoins. La reconnaissance vocale de Google est entraînée sur un grand nombre de mots et fournit une liste de mots par rapport à ce qui a été reconnu. Il est ensuite nécessaire de filtrer cette liste de mots en fonction du contexte de l'application. Ce traitement sur la liste peut être effectué lors de l'apprentissage du module de reconnaissance vocale ou par un module déjà

entraîné et non modifiable. Ce traitement sera effectué après la sortie pour exclure les mots non souhaités et identifier le mot adéquat à la situation.

1.2.3.2 Module d’auto correction

Tel que présenté précédemment, le module de reconnaissance vocale propose pour chaque mot reçu une liste de mots pour lesquels correspondent des calculs probabilistes [44] (ex : entrée \rightarrow "dix heures", sortie \rightarrow ["10h", "dix heures", "Deezer"]). Le système de reconnaissance vocale comprend qu’il est probable que l’utilisateur puisse parler d’heure ou du site de musique, car la prononciation est proche). La sortie sélectionnée dépend de l’entraînement qu’a subi le module de reconnaissance vocale. Dans le cas où il n’est pas possible de modifier l’entraînement, la solution est de prendre en compte le contexte pour connaître la meilleure sortie à utiliser. Cela donne deux solutions :

- Entraîner un module de reconnaissance vocale pour ne reconnaître que les mots liés aux besoins souhaités ;
- Utiliser un module de reconnaissance vocale déjà entraîné et filtrer la sortie du module pour récupérer les mots correspondant au contexte d’utilisation.

En reprenant l’exemple précédent, si le contexte du dialogue est musical, le mot "Deezer" sera privilégié à "10h", et dans le cas d’un contexte de prise de rendez-vous, ça sera "10 heures".

1.2.3.3 Les tours de paroles

Dans les premiers systèmes de dialogue homme-machine (DHM), il était courant de voir un bouton permettant à l’utilisateur d’indiquer lorsqu’il commençait à parler et lorsqu’il finissait. Ce sont des systèmes " de commande et de contrôle" [45]. Cela oblige l’utilisateur à appuyer sur un bouton dès qu’il souhaite parler. Cette méthode est acceptable lors de requêtes isolées, mais cela devient problématique si l’utilisateur converse avec la machine. Aujourd’hui, il existe des systèmes qui s’occupent automatiquement de détecter le début et la fin d’une intervention, les systèmes "conversationnels" [46]. Néanmoins, les systèmes détectant automatiquement la fin d’une phrase ont aussi leurs inconvénients. Comment être sûr que l’utilisateur a réellement terminé de

1.2. IMPORTANCE DU DIALOGUE HOMME-MACHINE

parler ? Par exemple, les utilisateurs fragmentent parfois leurs phrases. La personne dit "Je veux..." se met à réfléchir à la suite de la phrase perdant trop de temps. Ceci provoque l'arrêt de l'écoute du système, qui analyse la phrase reçue qui est incomplète. Selon l'utilisation du vocal, une solution est plus adaptée qu'une autre. Dans le cas d'une détection automatique de fin de phrase, il est possible de jouer avec le temps d'écoute. Cela est à adapter selon le public visé. Pour un public plus âgé, il est possible de rallonger le temps d'écoute pour pallier au problème de fragmentation de phrase qui est fréquent. Mais, le temps d'écoute de la machine ne doit pas être trop long, pour éviter que l'utilisateur pense que la machine ne l'écoute pas, car elle ne réagit pas, alors qu'elle est encore en train d'écouter. Une solution peut être d'avoir un visuel montrant l'état de la machine, comme trois petits points ou un cercle de chargement pour montrer qu'une action de la machine est en cours.

Les trois caractéristiques présentées montrent ce qui entoure le dialogue. Il faut maintenant concevoir ce dialogue.

1.2.4 Étapes du design du dialogue

Pour introduire le design d'Interaction Utilisateur Vocale (IUV), il faut introduire les systèmes à réponse vocale interactive (RVI) [38] car ils représentent l'apparition des IUV. Les systèmes RVI sont utilisés pour aider l'utilisateur à choisir certaines options. Cela permet à l'appelant de pouvoir obtenir une réponse à sa question sans être directement en relation avec une personne. Cela offre l'avantage de prendre son temps pour répondre. Cependant, ces systèmes, efficaces dans leurs cas d'utilisation, ne sont pas suffisamment poussés pour répondre aux attentes du projet. Pour le projet, l'utilisateur peut faire des demandes variées qui ne suivent pas toujours un ordre prédéfini. Le but est de garder un aspect naturel. De plus, la réalisation d'un système de dialogue homme-machine (DHM) est complexe si l'on veut aller au-delà des simples suites de questions-réponses [47]. Avant de se lancer dans la mise en place du dialogue, certaines questions se posent :

- Comment représenter le dialogue ?
- Quelle est la profondeur du dialogue recherchée ?
- Comment rendre le dialogue compréhensible ?

1.2.4.1 Un aspect visuel

L’aspect visuel du dialogue peut avoir lieu sous deux formes qui peuvent très bien cohabiter :

- Une visualisation textuelle du dialogue (chatbot [48] ou interface graphique) ;
- Un agent virtuel.

Chatbot ou interface graphique

Un chatbot est un “programme visant à stimuler une conversation avec les utilisateurs” [49]. Dan Grover stipule que : "les bots ne remplaceront jamais des applications. Selon applications remplaceront les anciennes." [13]. Il y compare deux bots démontrant qu’une interface graphique (GUI) liée à la discussion en cours avec le chatbot apporte une meilleure compréhension de l’application pour l’utilisateur. Cependant, comme le dit Emmet Connolly "les bots doivent être utilisés dans le but d’améliorer l’expérience utilisateur." [50]. Le fait d’avoir un dialogue écrit permet, dans le cas d’une utilisation par une population âgée, de garder le fil de la conversation et de confirmer si le bot a compris ou non ce que l’utilisateur vient de dire [51].

Agent virtuel

Un agent virtuel offre un visage sur la voix que l’utilisateur entend, créant ainsi un lien de confiance et diminuant l’anxiété face à une voix inconnue [52]. L’agent permet de rendre une interaction plus humaine [53]. L’agent virtuel n’est cependant pas toujours la solution. L’utilisateur peut se sentir observé [54] et préfère un simple système de dialogue. Il est donc nécessaire de déterminer le public ciblé et de laisser le choix aux utilisateurs l’usage d’un agent virtuel, étant donné qu’un système vocal simple peut suffire. Mais l’agent virtuel peut rendre le dialogue plus vivant, entre autres par des gestes qui indiquent à l’utilisateur quand prendre la parole et quand écouter [55].

1.2.4.2 Profondeur du dialogue

Comme évoqué précédemment, le dialogue peut ne consister qu’en de simples commandes vocales, mais peut aussi être une suite de questions/réponses. De nouveau, la typologie du dialogue est impliquée. En effet, un dialogue maître (utilisateur) - maître (système) est plus poussé, car chaque phrase est une nouvelle phrase en réponse à

1.2. IMPORTANCE DU DIALOGUE HOMME-MACHINE

une autre, tandis que dans le cas d'un dialogue maître (utilisateur) - esclave (système), le système réagit de manière prédéfinie. L'utilisateur peut utiliser différentes formulations mais le système réagit de la même manière et fournit les mêmes réponses. Le tableau 1.1 montre deux exemples de dialogue. Le premier est un dialogue maître-maître et le deuxième est un dialogue maître-esclave.

tableau 1.1 – Exemple d'un dialogue maître-maître et d'un dialogue maître-esclave

Maître-Maître	Maître-Esclave
Exemple 1	Exemple 2
Utilisateur (U): J'apprécie aller naviguer en mer.	U: Je souhaite ajouter un rendez-vous.
Système (S): Il paraît que l'air de la mer est rafraichissant.	S: Très bien, à quelle date dois-je mettre le rendez-vous?
U: Oui, mais il faut être prudent en jour de tempête!	U: Je souhaite que tu le mettes pour demain s'il te plaît.
S: En consultant les informations, j'ai noté qu'un bateau s'est échoué sur la plage l'autre jour.	S: C'est noté et quelle sera l'heure du rendez-vous?

L'exemple 1 montre que le dialogue maître-maître est plus profond que dans l'exemple 2, où le dialogue consiste en une suite de questions-réponses où les réponses sont prédéfinies pour réagir à certaines questions. Le dialogue maître-esclave de l'exemple 2 peut donc être autant élaboré que le dialogue maître-maître au niveau du langage et de la syntaxe, mais il est moins profond. Cela amène donc le point sur l'élaboration du dialogue et comment le rendre compréhensible.

1.2.4.3 Créer un dialogue compréhensible

Pour que le dialogue soit compréhensible par l'utilisateur, il faut suivre certaines règles de conception.

1.2.4.3.1 Un dialogue facile à suivre Dans un dialogue, une réponse est généralement liée à une demande. Lorsqu'un sujet est en cours, il ne faut pas chercher à changer de sujet [56]. Effectivement, si la conversation n'a pas de structure logique, l'utilisateur se sent perdu. Cependant, il arrive que des changements de sujets surviennent, soit pour revenir par la suite au sujet initial, soit pour entraîner la discussion sur un nouveau sujet si cela ne bloque la discussion qui était en cours. Le tableau

1.2 montre trois exemples de conversations et montre l'importance de la structure du dialogue.

tableau 1.2 – Exemple de discussions avec et sans changement de sujet

Exemple 1	Exemple 2	Exemple 3
Utilisateur (U): Je souhaite ajouter un rendez-vous.	U: Je souhaite ajouter un rendez-vous.	U: J'aime bien regarder les bateaux.
Système (S): Quelle est la date du rendez-vous?	S: Quelle est la date du rendez-vous?	S: Je trouve que les voiliers sont les plus intéressants à contempler.
U: Jeudi prochain.	U: Quelle est la météo d'aujourd'hui?	U: Au fait, as-tu entendu parler du dernier film qui est sorti?
S: C'est noté, à quelle heure dois-je mettre le rendez-vous?	S: Il fait 10 °C est le ciel est dégagé. Voulez-vous continuer l'ajout de rendez-vous?	S: Quel est le titre du film?

Dans l'exemple 1, il n'y a pas de changement de sujet, tout se déroule normalement. Dans l'exemple 2, l'utilisateur change de sujet alors qu'il est en pleine action. Le dialogue est ici maître-esclave. Le système répond à la demande de l'utilisateur bien que ce dernier n'ait pas terminé l'action sur le rendez-vous en cours. Le système reprend ensuite la conversation qui était en cours, ce qui permet de recentrer l'utilisateur dans la conversation. Dans l'exemple 3, le dialogue est ici maître-maître. L'utilisateur et le système discutent de sujets divers. L'utilisateur a beau changer de sujet, cela ne mène pas à une incompréhension. Les sujets de conversation étant libres, il est possible de changer de sujet et d'en débiter un nouveau.

Chaque type de dialogue a donc un mode de traitement qui lui est propre. Cependant, les dialogues utilisés doivent avoir en commun un niveau de langage adapté au public visé.

1.2.4.3.2 Un niveau de langage adapté Il existe trois niveaux de langage : le langage familier, le langage courant et le langage soutenu. Le niveau de langage est choisi selon le public visé [57]. Par exemple, des jeunes d'une dizaine d'années n'auront pas le même niveau de langage que des aînés ayant 70 ans et plus. Cependant, il faut maintenir un dialogue simple et naturel, quel que soit le niveau de langage à adopter [58]. De plus, pour un même niveau de langage, les expressions peuvent différer selon les générations.

1.2. IMPORTANCE DU DIALOGUE HOMME-MACHINE

1.2.4.3.3 Diversifier le dialogue Même en adoptant un niveau de langage un dialogue doit être diversifié. S'il est demandé à plusieurs personnes de formuler une phrase sur une idée, celles-ci exprimeront différemment cette idée. Par exemple, les demandes de consultations de la météo sont variables. "Je veux connaître la météo", "Quelle est la météo de demain?", "Quel temps fait-il.".... Avoir un dialogue diversifié ne change pas la compréhension du dialogue, mais l'améliore en donnant un aspect plus humain. Si une machine répète en boucle les mêmes réponses, l'aspect d'un dialogue plus naturel entre l'utilisateur et la machine disparaît. Des études psychologiques sont réalisées pour améliorer cet aspect de relation naturelle entre l'homme et la machine [59].

1.2.4.3.4 Enrichir le dialogue En plus de diversifier le langage il est possible de l'enrichir avec la voix répondant à l'utilisateur. Les personnages utilisent leur voix pour exprimer leurs sentiments [60]. De plus, les personnes sont plus réceptives à certaines voix, des voix féminines, des voix masculines, graves ou aiguës [61]. En effet, la voix est importante car des malentendus peuvent survenir dans les échanges non verbaux, car il n'y a pas d'intonation [62]. Les émoticônes sont parfois utilisés pour pallier à la neutralité de l'écrit. Le dialogue doit donc refléter la richesse de l'échange sonore. Il est donc important de choisir une certaine intonation selon "l'atmosphère" du dialogue, de même que les mots utilisés.

Maintenant que les informations nécessaires à la mise en place du dialogue sont réunies, il faut structurer le dialogue.

1.2.5 Avoir un dialogue structuré

Pour chaque type de dialogue, une structure doit être adoptée. En effet, un dialogue maître-esclave a pour but d'aider à répondre aux demandes de l'utilisateur, tandis qu'un dialogue maître-maître est un échange d'égal à égal entre l'utilisateur et la machine. Par exemple, pour instaurer un dialogue maître-esclave, il faut récupérer de l'information des demandes de l'utilisateur. On parle alors d'une solution par reconnaissance de forme ou "pattern matching" [63]. Plusieurs langages d'intelligence artificielle proposent cette solution, comme AIML, Rivescript ou ChatScript. Il est alors nécessaire de concevoir une structure pour le dialogue afin de récupérer

des informations. Une pratique pour concevoir la base de cette structure est d'écrire des exemples de dialogue [64]. En écrivant ces exemples, cela permet de savoir les sujets qui doivent être retrouvés dans le dialogue. La figure 1.1 montre l'exemple de la structure d'un dialogue autour des rendez-vous.

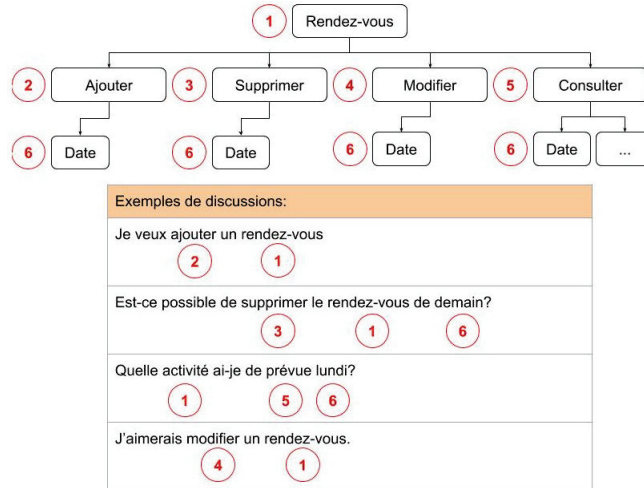


figure 1.1 – Exemple de la structure du dialogue autour de la prise de rendez-vous

Les exemples de discussion permettent de définir des termes communs et de créer des sous-structures plus spécifiques. La structure du dialogue permet de traiter chaque demande de manière adaptée et donc de confirmer à l'utilisateur que sa requête a été reçue.

1.2.6 Retour du système sur la demande de l'utilisateur

La confirmation ou l'infirmité de la demande de l'utilisateur est une caractéristique du DHM [65]. Le retour peut soit être une confirmation du système à l'utilisateur pour dire que la demande de l'utilisateur est comprise, soit une demande de précision ou de reformulation car le système n'a pas compris l'utilisateur. Il existe différents types de confirmation listés ci-dessous :

- Les tiers de confiance [66] [67];
- Les confirmations implicites [68];
- Les confirmations génériques [69];

1.2. IMPORTANCE DU DIALOGUE HOMME-MACHINE

- Les confirmations visuelles [70].

Tiers de confiance

Le tiers de confiance se base sur le pourcentage de compréhension. Comme le dialogue fonctionnant par correspondance de motif ou "pattern matching", il est possible d'évaluer le taux de compréhension et de donner un retour selon trois cas possibles.

- Compréhension supérieure à 80%.

Ex : U : "Je veux connaître la météo du jour."

S : "Il fait ensoleillé."

La phrase est totalement reconnue, donc le système répond à la demande.

- Compréhension entre 79 et 45%.

Ex : U : "Je veux tendre un rendez-vous"

S : "Vouliez-vous prendre un rendez-vous?"

Le système se rend compte que "tendre" ne fait pas de sens avec le reste de la phrase mais suppose avec le contexte de rendez-vous que "tendre" a pu être confondu avec "prendre".

- Compréhension inférieure à 45%.

Ex : U : "J'aime les bateaux." En imaginant qu'aucune règle ne prend en compte un sujet parlant de bateau, le système répond alors :

S : "Je n'ai pas compris."

Le système n'est pas conçu pour traiter ce cas et ne connaît pas de contexte proche pour supposer qu'il y a pu avoir une mauvaise formulation. Le système se contente d'expliquer qu'il n'a pas compris la demande.

Ce type de confirmation peut être jumelé avec d'autres confirmations.

Confirmations implicites

Lors des confirmations implicites, l'utilisateur est avisé de ce qui a été compris en reprenant une partie de la phrase d'entrée. Cela n'implique pas que toute la demande a été comprise. En reprenant l'exemple précédent :

U : "Je veux tendre un rendez-vous"

S : "Vouliez-vous prendre un rendez-vous?"

La réponse indique à l'utilisateur que le système a constaté qu'il parle de rendez-vous en reprenant ce terme, et indique également de ce qui ne peut pas être compris.

Confirmations génériques

Les confirmations génériques s'appliquent plus facilement dans le cas d'un dialogue libre, puisqu'elles visent à ouvrir le dialogue. Le tableau 1.3 donne l'exemple d'une confirmation implicite qui n'apporte pas un aspect naturel dans la conversation. A l'inverse, la confirmation générique, qui est à la fois une réponse naturelle, ouvre au dialogue.

tableau 1.3 – Comparaison d'une confirmation implicite et d'une confirmation générique

Confirmation implicite	Confirmation générique
S: "As-tu passé une bonne journée?"	S: "As-tu passé une bonne journée?"
U: "Oui ça va."	U: "Oui ça va."
S: "Je suis content que tu es passé une bonne journée."	S: "Oh dis moi ce qu'il s'est passé durant ta journée."

Dans les deux dialogues, l'utilisateur reçoit un retour confirmant qu'il est compris, mais la seconde confirmation est plus naturelle selon l'aspect humain.

Les confirmations visuelles

Il n'a été évoqué jusqu'ici que des confirmations vocales, or une interface graphique peut apporter ainsi une bonne compréhension [13]. Dans le cas d'une application incluant une GUI, il est possible de signaler à l'utilisateur qu'il a été compris en réalisant sa requête visuellement. En prenant l'exemple d'un répertoire, si l'utilisateur demande :

U : "Je souhaite ajouter le contact Hugo à mon répertoire"


Il est intéressant de montrer visuellement une fiche de contact avec le prénom Hugo et de mettre à jour cette fiche à chaque nouvelle information reçue. La figure 1.2 schématise un exemple de retour visuel sur la conversation entre un agent virtuel et l'utilisateur pour le remplissage d'une fiche contact.

1.2.6.1 Difficultés liées aux interactions vocales

Le but du DHM est de parler vocalement avec la machine et non de parler en écrivant le texte à la main. Cela amène des difficultés puisqu'il faut reconnaître ce que l'utilisateur dit.

Reconnaissance vocale

1.2. IMPORTANCE DU DIALOGUE HOMME-MACHINE



U: Je souhaite ajouter le contact Hugo à mon répertoire

S: Très bien, quelle autre information dois-je enregistrer?

U: Sa date de naissance est le 15 juin 1989

Fiche contact	
Prénom: Hugo	Nom:
Date de naissance: 15 Juin 1989	
Adresse:	
Numéro de téléphone:	
Adresse courriel:	

figure 1.2 – Exemple de retours visuels lors de l'ajout d'un contact

CHAPITRE 1. REVUE DE LITTÉRATURE

La reconnaissance vocale automatique pose des problèmes lorsqu'on l'utilise dans un DHM. La reconnaissance vocale est le passage d'un signal audio à une transcription suivant un code plus ou moins proche du langage écrit. Ceci nécessite une grande source de données. Il existe aujourd'hui des systèmes très développés contenant un très grand nombre de données, tels que Google Now ou Siri [71]. Ceux-ci offrent des taux de reconnaissance vocale élevés. Mais certains mots sont parfois confondus avec d'autres, dû à une proximité dans leur prononciation. De plus, de nombreux noms propres sont mal reconnus si le système ne connaît pas ces mots, comme les noms de rues. Il est donc possible de corriger ce problème en modifiant la base de connaissances du système de reconnaissance [72]. Cependant, certains systèmes libres d'accès déjà entraînés donnent de bons résultats mais n'autorisent pas la personnalisation, tandis que les systèmes libres permettant des modifications ne sont pas ou peu entraînés.

L'utilisation de plusieurs systèmes de dialogue

Le paragraphe sur la typologie montre les différents dialogues pouvant exister. Pour pouvoir traiter ces types de dialogue, il est préférable d'utiliser des solutions différentes. Néanmoins, cela apporte une difficulté qui est de savoir quand utiliser les modèles. En effet, le choix du modèle dépend de l'utilisation qui doit être faite pour l'application. Par exemple, pour un agenda qui nécessite la prise de rendez-vous, le dialogue par reconnaissance de motifs permet de recueillir les informations données par l'utilisateur et de diriger le dialogue pour prendre un rendez-vous. Mais si l'application doit permettre également à l'utilisateur de converser avec la machine sur des sujets plus généraux, un deuxième modèle de dialogue plus général doit être mis en place. Cependant, à tout moment il peut être préférable de privilégier l'utilisation du dialogue de reconnaissance de motifs pour analyser la phrase de l'utilisateur et de ne faire intervenir le second modèle que si le premier n'a pas pu donner de réponse, comme illustré dans la figure 1.3. Les différents aspects auxquels il faut réfléchir pour mettre en place un dialogue ont été évoqués. Il faut maintenant développer les solutions existantes pour construire un DHM.

1.3. DIFFÉRENTES SOLUTIONS POUR LE DHM

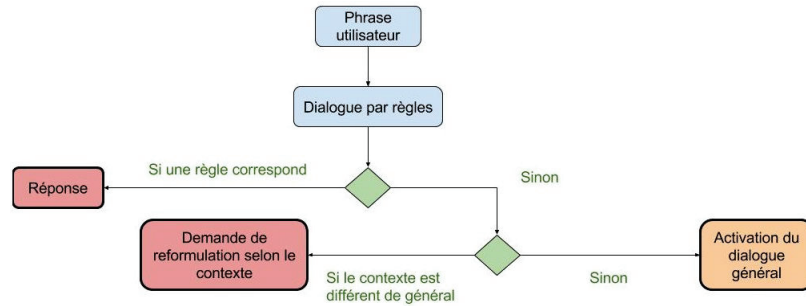


figure 1.3 – Cohabitation de deux types de dialogue

1.3 Différentes solutions pour le DHM

Le DHM a vu ses débuts dans les années 60 avec ELIZA [73], qui ne comprenait que 53 règles pour la reconnaissance de motifs et qui donnait l'illusion d'interagir avec l'utilisateur en substituant une partie de la phrase de l'utilisateur pour la transformer en question. Le fait de poser beaucoup de questions permettait de cacher l'ignorance du programme. ELIZA a aujourd'hui évolué en de nombreux chatbots utilisant désormais des dizaines de milliers de règles, tels que ChatScript, AIML et pleins d'autres. Les Chatbots ou bots, sont conçus pour simuler une conversation humaine. Ils peuvent utiliser des IUV mais sont principalement constitués d'un environnement d'interface textuel (TUI). Des études ont montré qu'une interface graphique est parfois plus efficace qu'un bot n'utilisant que du texte [13]. Maintenant, les chatbots intègrent généralement une interface graphique (GUI), en plus du texte, pour permettre une meilleure compréhension et de meilleures interactions. Il faut néanmoins garder à l'esprit qu'un bot n'est nécessaire que s'il apporte une aide à l'utilisateur au niveau de l'expérience avec la technologie. Une autre approche plus récente au DHM est l'utilisation de réseaux de neurones [74]. Il existe différents réseaux de neurones pour des traitements différents. Nous nous intéresserons ici au *sequence to sequence* [75]. Nous allons dans un premier temps parler des solutions au dialogue par reconnaissance de motifs, puis aux solutions pour un dialogue général.

1.3.1 Modèles à base de reconnaissance de motifs

L'idée du dialogue par reconnaissance de motifs ou également appelé dialogue par règles est de réaliser une action selon la demande de l'utilisateur. Le but est de pouvoir connaître le sujet/contexte pour traiter chaque demande de manière appropriée. On peut citer le modèle Latent Dirichlet allocation (LDA). Ce modèle découvre les sujets liés à un document. Le modèle LDA applique des probabilités à chaque mot pour ensuite en tirer les sujets auxquels ils sont liés. Il postule que les mots contiennent une information sémantique importante, ce qui implique que les phrases sur un sujet similaires utiliseront probablement des mots équivalents ou proches [76]. Cependant, ce modèle demande un apprentissage et donc des données pour réussir à reconnaître le contexte d'une phrase ou d'un document. Une autre approche est l'utilisation d'une correspondance de motif. Il est possible de trouver un grand nombre d'applications permettant de mettre en place un Chatbot construit autour de règles. Il existe différents bots en ligne avec lesquels on peut entamer une conversation, tels que Mitsuku [77] (ayant remporté le prix Loebner en 2016 [78]) ou Cleverbot [10], qui apprend des conversations avec les utilisateurs. Plusieurs logiciels construisent des bots à partir d'une structure par règles, et ce même pour des débutants. Par exemple, Mitsuku se base sur AIML [77] mais on peut aussi citer Chatscript [79] ou Rivescript [80]. Tous trois se basent sur une structure similaire utilisant un système de correspondance de motifs. L'idée est de créer plusieurs motifs pour traiter plusieurs entrées d'une manière différente en utilisant des systèmes de priorité, voir figure 1.4.

Dans cette figure, un système de priorité est instauré où le dièse a une priorité supérieure à l'accent circonflexe. Une phrase contenant le mot "bonjour" et ne contenant pas le mot "météo" activera la règle avec le terme "bonjour". Une phrase contenant le mot "météo" et ne contenant pas le mot "bonjour" activera la règle avec le terme "météo". Une phrase contenant le mot "bonjour" et le mot "météo" activera la règle avec le terme "météo", car la règle s'activant pour le terme "météo" est prioritaire.

En 2013 a été menée une étude sur des personnes atteintes de troubles cognitifs et des personnes âgées [6]. Cette étude se focalise sur la capacité de ces personnes à détecter les erreurs vocales lors d'un ajout de rendez-vous avec un agent virtuel, sur un calendrier interactif. Le dialogue se base sur des règles pour comprendre et répondre à l'utilisateur, et il est montré que ces personnes sont plus à l'aise à suivre

1.3. DIFFÉRENTES SOLUTIONS POUR LE DHM

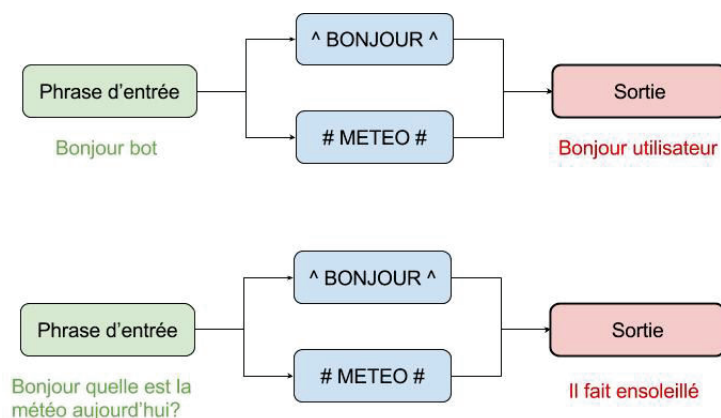


figure 1.4 – Système de règles et de priorités (ici dièse plus important que l’accent circonflexe)

l’agent virtuel et à détecter des erreurs lorsque les retours vocaux sont plus fréquents.

Il est donc possible de construire un dialogue autour de règles. Cependant, cette solution est peu adaptée pour un dialogue général.

1.3.2 Modèles pour un dialogue général

Cleverbot se développe grâce aux entrées des utilisateurs [10]. Cela permet d’avoir une discussion générale homme-machine. Cependant, une idée peut-être exprimée selon des manières différentes, en utilisant entre autre des synonymes ; Il faudrait donc ajouter une règle pour chaque nouveau cas qui n’est pas encore traité et à chaque fois il faudrait vérifier que la priorité d’un nouveau motif ne pose pas de problème avec un ancien. Construire un dialogue général à partir de règles n’est pas une solution appropriée, car il faudrait donc rajouter un nombre trop important de règles pour couvrir le plus de cas possibles. Il existe des solutions comportant des agents comme Siri, Alexa, ou Ok Google, qui sont des technologies permettant à un utilisateur de prendre des notes, d’allumer une télé, de faire des recherches [8]. Cependant, dès lors que l’utilisateur fait une demande plus large qui ne peut être traitée par l’agent, une recherche est lancée sur internet pour retourner les résultats correspondant à la demande de l’utilisateur. Cela n’engage pas dans un dialogue puisque l’utilisateur doit par la suite étudier lui-même les résultats qu’il a reçus.

Une autre approche pour ce type de dialogue est l'utilisation de réseaux de neurones [74]. Il existe différents types de réseaux, mais certains sont plus appropriés pour la construction d'un dialogue. On évoque principalement les "Retrieval Models" ou modèles d'extraction [81] et les "Generative Models" ou modèles génératifs [82]. Les modèles d'extraction sont aisés à comprendre mais ne génèrent pas de nouveau texte. Ce type de modèle sélectionne une réponse dans sa base de données pour répondre à la demande qui est faite. Cela est utile pour réaliser une tâche spécifique et répondre à l'utilisateur selon les mots clés de la requête. En comparaison, un modèle génératif est plus complexe mais permet de générer de nouvelles réponses et de s'adapter aux entrées de l'utilisateur.

1.3.2.1 Modèles de réseaux de neurones propres au langage

Dans les modèles de réseaux de neurones pour le langage, il existe les modèles n-grams, les modèles généralistes et les modèles de récupération (retrieval model). Les **modèles n-gram** ont pour but de prédire la suite d'un mot ou d'une phrase [83, 84]. Ces modèles donnent des probabilités à une phrase ou une séquence de mots sur le prochain mot à venir selon les mots déjà présents. Par exemple, en tapant un mot sur un smartphone, des propositions de mots apparaissent et changent à chaque nouvelle lettre rentrée. Cependant, cela est plus utilisé dans la correction orthographique ou encore la reconnaissance vocale et non pour la construction d'un dialogue. Un modèle génératif ou de récupération sont plus adaptés aux besoins pour un dialogue général, cependant le modèle génératif va créer du texte par rapport à un sujet sans répondre à une question et le modèle de récupération va quant à lui donner une réponse par rapport à une question sans créer du texte [85]. Un modèle de récupération courant dans cette pratique de création de dialogue est le modèle "sequence to sequence" (seq2seq) [75]. Ce modèle est constitué d'un "encodeur" et d'un "décodeur". Le but est de convertir une phrase en une réponse. Les modèles seq2seq sont beaucoup utilisés pour la traduction. Cependant, ce qui doit être mis en place est un dialogue. Pour mettre en place un dialogue, il faut effectuer un apprentissage à partir de données comprenant des questions-réponses, permettant ainsi au modèle d'apprendre de ces exemples pour ensuite générer de nouvelles phrases. Un des défis est de trouver une quantité de données suffisante pour l'apprentissage. En effet, plus

1.3. DIFFÉRENTES SOLUTIONS POUR LE DHM

le nombre de données sera important, plus le modèle diversifiera son apprentissage et développera ses connaissances.

Nous venons de parler des questions à se poser pour savoir comment créer un dialogue, ainsi que les solutions pour développer et mettre en place le dialogue. Nous allons maintenant développer les objectifs du projet et la méthodologie.

Chapitre 2

Objectifs et méthodologie

2.1 Contexte du calendrier Amelis

Avant ce projet Amelis, il existait déjà un calendrier propre aux personnes atteintes d'Alzheimer. Ce précédent projet porte le nom d'AMELIS, signifiant Alzheimer MÉmoire et LIens Sociaux. AMELIS implique une bonne compréhension des besoins évolutifs sur le plan cognitif et physique des personnes avec une démence. Il vise à utiliser la gérontechnologie pour que les personnes avec la maladie d'Alzheimer puissent maîtriser le fonctionnement d'un outil technologique. Le nouveau projet Amelis, quant à lui, porte le même nom mais sans les majuscules. Ce projet vise à aider à l'organisation de la vie quotidienne d'aînés, ayant ou non de faibles troubles cognitifs, à travers un calendrier interactif [86]. Amelis regroupe donc plusieurs buts que nous allons maintenant évoquer.

2.1.1 Buts généraux du calendrier Amelis

Le projet Amelis a vu le jour suite à des expérimentations chez des aînés. Il est constaté que ces personnes utilisaient soit des calendriers muraux souvent trop petits pour noter tous leurs rendez-vous, soit des post-it que les personnes âgées avaient du mal à retrouver. Le projet Amelis a pour but de :

- Permettre aux aînés de préserver une autonomie dans leur vie quotidienne. Les

2.1. CONTEXTE DU CALENDRIER AMELIS

aînés font face, au fil des ans, à plusieurs décisions difficiles, par exemple lors de la demande de services à domicile ou d'hébergement en résidence privée. Thomas Kleinberger et al. soulignent que le but d'une technologie ambiante permet d'aider les personnes, handicapées ou âgées, à vivre dans leur environnement préféré plus longtemps en répondant à des besoins spécifiques [1];

- Adapter les technologies aux aînés pour que leurs interactions soient simples et compréhensibles. Durant les dernières décennies, les technologies ont radicalement évolué. Depuis dix ans, les appareils sont devenus plus personnalisés. L'efficacité des applications et du design d'interfaces s'adressent aujourd'hui à une grande diversité de personnes [87], plus particulièrement auprès des personnes âgées, qui ont des besoins spécifiques dus au vieillissement ;
- Mettre en place un agent virtuel pour faciliter l'interaction entre les aînés et les technologies. Ramin Yaghoubzadeh développe les bénéfices de l'utilisation d'un agent virtuel et la manière de le concevoir à l'aide de groupes participatifs avec les aînés [6].
- Instaurer un dialogue convivial entre l'utilisateur et l'agent virtuel. Le dialogue a pour but d'offrir un mode d'interaction différent du tactile et d'alléger les tâches cognitives de l'utilisateur ;
- Intégrer des émotions à l'agent virtuel et étudier les émotions que le calendrier et l'agent virtuel créent chez les utilisateurs, pour rendre l'interaction plus humaine et plus conviviale, et faciliter l'approche des personnes âgées face à la technologie [88] [89].

Pour répondre à ces objectifs, le calendrier Amelis est proposé.

2.1.2 Étapes de conception du calendrier Amelis

Pour concevoir le calendrier, une méthodologie de conception participative est adoptée auprès de personnes âgées autonomes vivant dans des résidences au Québec et en France [9]. Trois ateliers de conception participative ont eu lieu :

- Le premier atelier, regroupant un total de 27 participants âgés vivant au Québec et en France, a pour but de connaître les fonctionnalités à intégrer au calendrier. Deux versions du calendrier sont ressorties, une version québécoise et

une version française. Ces versions sont très similaires à l’exception de certaines fonctionnalités, comme l’accès à la musique (figures A.1 et A.2 de l’annexe A). Pour éviter de créer deux versions, un prototype commun a été choisi, voir figure 2.1, qui se trouve également en annexe A.3 ;

- Le deuxième atelier, regroupant un total de 34 participants âgés vivant au Québec et en France, définit les modes d’interactions que les utilisateurs souhaitent trouver et les modes d’interaction à associer aux différentes fonctionnalités. Trois modes ont été proposés : tactile, vocal et agent virtuel.
- Le troisième atelier, regroupant un total de 14 participants âgés vivant au Québec et en France, est centré sur l’agent virtuel, et plus particulièrement son apparence. Le but est de définir l’aspect de cet agent, à savoir son sexe, son âge, la façon dont il est habillé, comment il s’exprime et dans quel cadre il doit se trouver (l’environnement qui se trouve derrière lui).

Suite à ces ateliers, les aînés ont manifesté leur intérêt pour les interactions avec un agent virtuel. Le but de ce mémoire est de concevoir l’interaction vocale de l’agent virtuel pour rendre plus convivial l’usage du calendrier. Les objectifs du mémoire et la méthodologie seront précisés aux paragraphes 2.3 et 2.4. Au préalable, le fonctionnement du calendrier Amelis va être présenté.

2.2 Fonctionnalités et aspect du calendrier Amelis

La conception participative avec les aînés montre l’importance de fournir des interactions qui facilitent l’usage du calendrier. L’aspect du calendrier Amelis, qui se trouve à la figure 2.1, permet de visualiser les fonctionnalités qui vont être présentées. Pour plus de clarté, le terme application Amelis servira à parler d’Amelis, car dans cette application, des fonctionnalités sont ajoutées à celles du calendrier, comme celle d’écouter de la musique. Lorsqu’on évoquera le terme calendrier, il sera associé à la page de l’application Amelis où se situe le calendrier. L’interface d’Amelis se divise en 3 parties :

Partie gauche :

La partie gauche de l’application Amelis est utilisée pour la navigation entre les pages et à chaque page est associ un nom et une icône. Ces icônes, ainsi que les

2.2. FONCTIONNALITÉS ET ASPECT DU CALENDRIER AMELIS

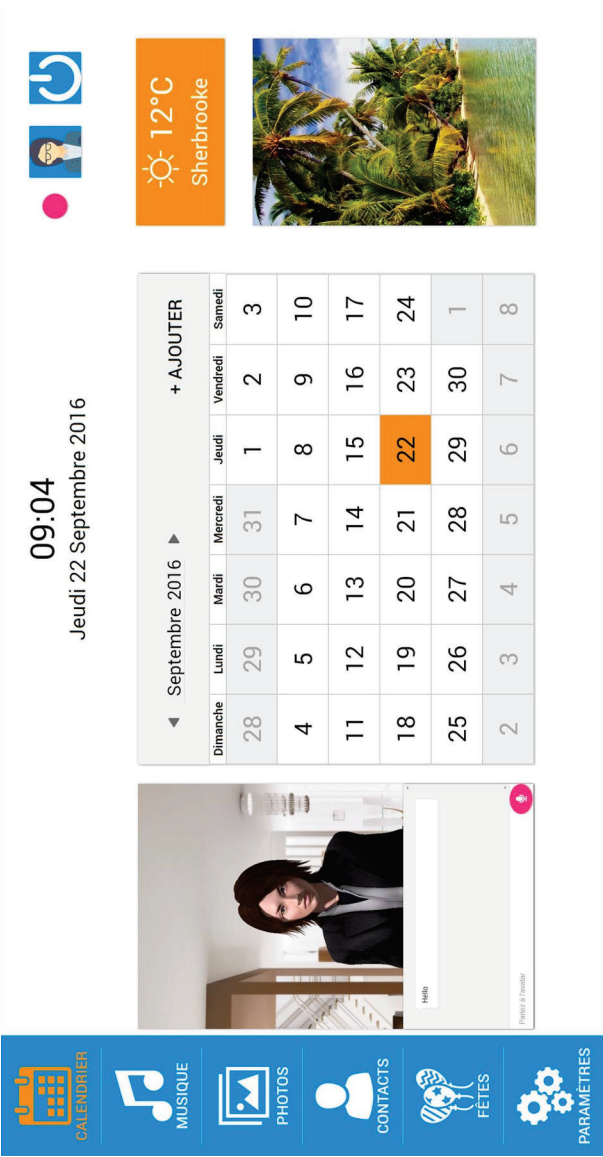


figure 2.1 – Maquette commune du calendrier Amelis entre le Québec et la France

CHAPITRE 2. OBJECTIFS ET MÉTHODOLOGIE

fonctionnalités de chaque page ont été choisies lors de la conception participative.

Partie droite :

La partie droite de l'application Amelis n'est visible que lorsque l'utilisateur se trouve sur la page du calendrier. Cette partie affiche la météo du jour, qui est un élément important que les participants souhaitaient retrouver, et une photo. Cette photo permet, suite à un clic dessus, de cacher l'agenda journalier et d'afficher la photo en grand. Lors de la conception participative, certaines personnes ont indiqué le point qu'elles ne voulaient pas que l'agenda, qui affiche les rendez-vous, soit toujours visible. En effet, elles préféraient garder une certaine intimité si des personnes venaient les visiter.

Dans le haut droit de l'écran se trouvent 3 boutons. Un point rouge à gauche sert à activer ou à désactiver l'enregistrement vocal. Le bouton central avec un personnage permet l'activation ou la désactivation de l'agent virtuel. Enfin, le bouton à droite permet d'éteindre l'écran. Ce dernier bouton a été mis en place pour éviter que l'utilisateur n'éteigne la tablette et qu'il faille redémarrer l'application. Certaines personnes âgées souhaitaient éteindre la tablette, soit pour diminuer la luminosité la nuit, soit pour mettre en veille l'application dans la journée.

Partie centrale :

La partie centrale varie selon la page. La page de l'agenda est différenciée des autres pages. L'agent virtuel ne peut apparaître que dans la page de l'agenda, car les fonctionnalités vocales sont centrées sur interactions avec l'agenda. Lorsque l'agent virtuel s'active, l'agenda rétrécit pour afficher l'agent virtuel et la conversation avec ce dernier. Cela permet à l'utilisateur de voir l'agent virtuel et de suivre la conversation, tout en continuant de pouvoir interagir avec l'agenda. Les autres pages ne contiennent ni agent virtuel, ni météo, ni photo d'accueil. La page de la musique affiche une liste de morceaux que l'utilisateur peut écouter. La page des photos regroupe plusieurs photos que l'utilisateur peut consulter et il peut définir la photo de son choix comme photo d'accueil. La page des contacts permet à l'utilisateur de consulter, modifier, ajouter ou supprimer un contact. La page des paramètres donne la possibilité à l'utilisateur de modifier son nom et celui de l'agent virtuel.

Cela résume les fonctionnalités se trouvant dans l'application Amelis. Toutes les fonctionnalités sont accessibles de manière tactile. Le vocal, quant à lui, est princi-

2.3. OBJECTIFS DE CETTE RECHERCHE

palement centré sur les rendez-vous, sur les demandes d'accès aux différentes fonctionnalités du calendrier, sur la navigation, sur la date et sur la météo. Ce qui nous amène aux objectifs de cette recherche.

2.3 Objectifs de cette recherche

L'objectif principal de cette recherche est de **fournir un dialogue entre une personne et un agent virtuel dans le cadre de la commande d'un calendrier**. Le dialogue doit s'intégrer dans l'utilisation globale de l'application Amelis avec l'agent virtuel, les commandes tactiles et l'aspect de l'agent virtuel. Deux contextes du dialogue ont été établis pour offrir une interaction conviviale, un premier contexte pour commander l'application Amelis et un deuxième contexte pour instaurer un dialogue libre. Plusieurs sous-objectifs, listés ci-dessous, ont été mis en place :

1. Construire une architecture adaptable en intégrant un dialogue et un agent virtuel dans l'architecture d'Amelis ;
2. Offrir un contrôle de l'application Amelis par la parole. Cela signifie que l'utilisateur peut activer des fonctionnalités de l'application Amelis en dialoguant. On parle ici de dialogue et non de commande par liste de mots comme dans le cas d'un RVI décrit dans le chapitre 1 [38];
3. Offrir des contextes de discussion libre entre l'agent virtuel et l'utilisateur pour donner un aspect plus humain lors des échanges ;
4. Intégrer un mode d'interaction tactile et un mode d'interaction vocal pour offrir un choix d'interaction plus large et plus simple à l'utilisateur ;
5. Évaluer les performances techniques et l'impact sur l'utilisateur du dialogue de contrôle et libre.

De chaque sous-objectif découle des hypothèses. Elles vont maintenant être décrites suivant chaque sous-objectif.

Hypothèses du sous-objectif 2 : offrir un contrôle de l'application Amelis par la parole : Pour répondre à ce sous-objectif, nous supposons qu'une analyse par un dialogue par règles permet de contrôler certaines fonctionnalités d'Amelis, dont la commande de l'accueil et les fonctionnalités liées à l'agenda.

1. Un dialogue par règles permet d'extraire les informations nécessaires pour contrôler une partie de l'application Amelis et de générer les réponses de l'agent virtuel. Les règles collectent les informations de la demande de l'utilisateur pour diriger l'application Amelis et pour créer des réponses ;
2. L'agent virtuel est capable avec le dialogue par règles de commander l'application Amelis avec diverses formulations. Les règles donnent la possibilité à l'utilisateur de faire une demande sans se soucier de la manière dont il la formule. Cela lui permet de parler librement sans être contraint à des commandes. De plus, ces règles permettent de récupérer les informations clés permettant de contrôler l'application Amelis en s'appuyant sur la demande de l'utilisateur ;
3. L'utilisateur est capable de comprendre les demandes de l'agent virtuel pour y répondre. Lorsque l'agent virtuel fait une demande précise, par exemple la demande du titre d'un rendez-vous, l'utilisateur est en capacité de comprendre cette demande et d'y répondre.

Hypothèses du sous-objectif 3 : offrir des contextes de discussion libre entre l'agent virtuel et l'utilisateur : Le dialogue libre offre une aisance de communication entre l'agent virtuel et l'utilisateur. Il a pour but de rendre les interactions plus conviviales lorsque l'utilisateur n'est pas en mode commande.

1. Le dialogue libre nécessite d'utiliser les réseaux de neurones pour le modéliser. Le dialogue par règles ne permet pas ou difficilement de mettre en place un dialogue libre ;
2. La qualité du dialogue libre impacte sur l'acceptabilité du dialogue par règles avec l'agent virtuel. L'hypothèse spécifie que l'absence d'un dialogue libre crée un manque au niveau du dialogue entre l'agent virtuel et l'utilisateur, mais la présence du dialogue libre influence, selon la qualité de ce dialogue, l'avis de l'utilisateur sur le dialogue qu'il entretient avec l'agent virtuel.

Hypothèses du sous-objectif 4 : intégrer un mode vocal et un mode tactile d'interaction :

1. Deux modes d'interaction élargissent les possibilités de choix d'interaction. L'utilisateur peut adapter ses actions selon le mode d'interaction qui lui convient et en changer quand il le désire ;

2.4. MÉTHODOLOGIE

2. La synchronisation entre les deux modes d'interactions simplifie leur cohabitation. Pour que l'utilisateur puisse change d'un mode à l'autre, les deux modes doivent communiquer.

Hypothèses du sous-objectif 5 : évaluer les dialogues : Les sous-objectifs précédents concernaient la conception et l'implantation du dialogue. Ce sous-objectif, qui concerne l'évaluation, se divise en 3 hypothèses montrant comment évaluer Amelis.

1. L'utilisateur est correctement compris par le modèle de reconnaissance vocale. Le module de reconnaissance vocale comprend les mots qu'énoncent des aînés québécois et français ;
2. Des tests techniques permettent de valider le bon fonctionnement du dialogue de l'application Amelis. Il est nécessaire de tester de façon technique toute l'application Amelis pour s'assurer de fournir une application Amelis fonctionnelle ;
3. Les utilisateurs apprécient le dialogue avec l'agent virtuel et trouvent plus facile l'utilisation de l'application Amelis avec l'agent virtuel. Il est supposé que ce dialogue est plus chaleureux et plus simple de communiquer de manière libre et plus humaine avec un agent virtuel, que ne n'avoir qu'un mode d'interaction tactile.

Maintenant que les objectifs et les hypothèses de cette recherche sont expliqués, le développement de la méthodologie va permettre de présenter le cheminement mis en place pour répondre à ces objectifs.

2.4 Méthodologie

La méthodologie suit un déroulement en quatre parties correspondant chacune à un objectif cité précédemment.

2.4.1 Utiliser des règles pour contrôler l'application Amelis et générer des réponses pour l'agent virtuel

La mise en place de règles doit suivre plusieurs étapes, qui vont être décrites par la suite, pour avoir un contrôle sûr et simple du calendrier et pour permettre à l'agent

virtuel de donner des réponses sensées et humaines. Les règles sont appliquées lors de l'utilisation des fonctionnalités d'Amelis, telles que les commandes de l'accueil et de l'agenda.

2.4.1.1 Mettre en place une structure pour le dialogue par règles

La structure du dialogue par règles concerne à la fois les règles permettant de récupérer les informations de l'utilisateur et les règles à donner à l'agent virtuel pour répondre à l'utilisateur.

Structure de récupération des informations de l'utilisateur : Les fonctionnalités tactiles étant déjà conçues, cela permet de connaître l'ensemble d'informations clés pour la création de la structure du vocal. Le dialogue par règles fonctionnant sur un principe de correspondance de termes, la structure des règles doit être basée autour de ces termes. Ceci impose l'écriture de conversations, afin d'identifier différentes tournures de phrases. Ceci va servir à créer différentes règles autour de termes récurrents. La figure 2.2 résume les étapes à suivre.

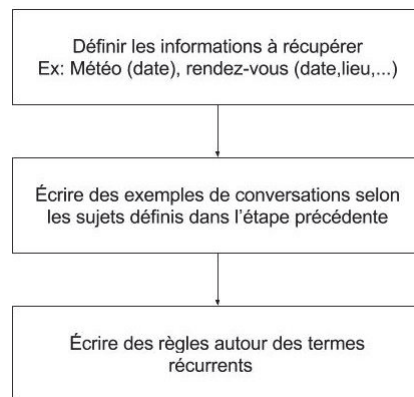


figure 2.2 – Étapes pour créer une structure pour les règles pour récupérer les informations de l'utilisateur

Structure des réponses à envoyer à l'agent virtuel : L'écriture d'exemples de dialogues nous oblige à réfléchir sur les réponses à donner selon les informations reçues. la structure de récupération des informations suit la méthodologie précédente. Les exemples de conversation vont permettre de générer les réponses de l'agent virtuel.

2.4. MÉTHODOLOGIE

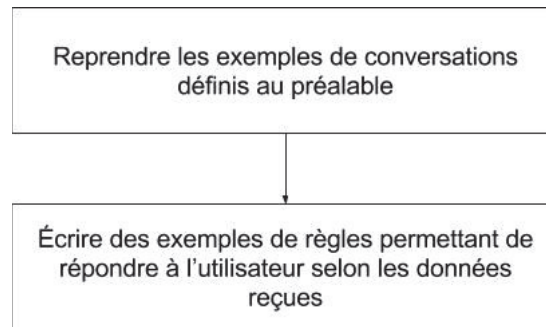


figure 2.3 – Étapes pour créer une structure pour les règles pour fournir des réponses à l'utilisateur

Une fois la structure définie et des exemples de règles écrits, comme résumé dans la figure 2.3, un vocabulaire doit être choisi. En effet, la première étape montre trop peu de variété de commandes du calendrier. Cela nécessite d'agrandir le vocabulaire, en particulier en rencontrant le public visé.

2.4.1.2 Adopter un vocabulaire diversifié pour comprendre et se faire comprendre par l'utilisateur

Des rencontres avec le public visé, ici des aînés, donnent la possibilité de discuter du type de langage à utiliser (familier, courant, soutenu) et de choisir si le vouvoiement s'impose. Pour diversifier le vocabulaire, l'écriture d'exemple des dialogues s'impose encore, cependant, le résultat s'améliore en demandant au public visé de réaliser cet exercice. L'écriture des règles se fait ensuite en utilisant les différentes tournures de phrases possibles qui vont suivre la structure définie au préalable. La figure 2.4 résume les étapes à suivre pour enrichir le vocabulaire à utiliser pour les règles.

La diversification du vocabulaire permet à l'agent virtuel de comprendre les diverses formulations de l'utilisateur. Cependant, les règles suivent un enchaînement précis. Pour que l'agent virtuel s'adapte à l'utilisateur, il faut refactoriser les règles pour agrandir les possibilités d'actions vocales, ce que la structure initiale de l'interaction vocale ne permet pas.

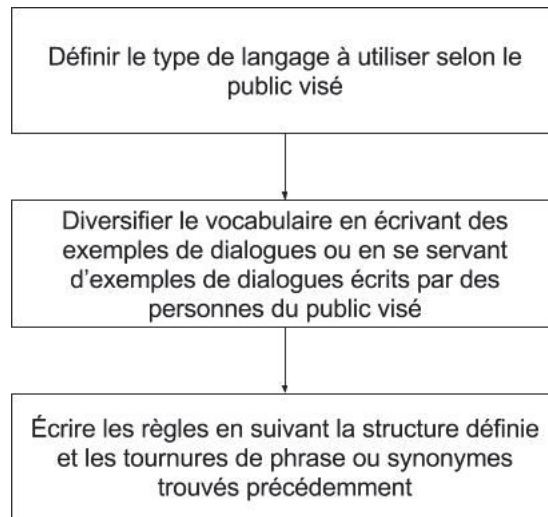


figure 2.4 – Étapes pour adopter le dialogue à utiliser pour les règles

2.4.1.3 Diversifier les interactions vocales pour simplifier les choix d'actions

Une règle est associée à chaque demande de l'utilisateur. Néanmoins, une demande peut comporter des sous-demandes. Par exemple, la demande pour connaître la météo "Je veux connaître la météo qu'il fera après-demain à Québec" intègre en plus une date et un lieu. En définissant les cas de demandes "multiples", la structure des règles se développe pour permettre l'imbrication de règles. L'agent virtuel s'adapte ainsi à l'utilisateur. La figure 2.5 résume les étapes à suivre pour simplifier les actions de l'utilisateur en diversifiant les interactions vocales.

Le dialogue par règles ne permettant que de contrôler l'application Amelis, une nouvelle alternative permet de mettre en place un dialogue libre avec l'agent virtuel au sein de l'application Amelis.

2.4.2 Utiliser un réseau de neurones pour un dialogue libre

L'utilisation de règles montre ses limites lors de dialogue sur des thèmes plus divers. Il faut donc utiliser une autre technique, telle que les réseaux de neurones. Le réseau de neurones doit être adapté à la création d'un dialogue et nécessite différents paramètres pour obtenir le dialogue libre attendu pour l'application Amelis.

2.4. MÉTHODOLOGIE

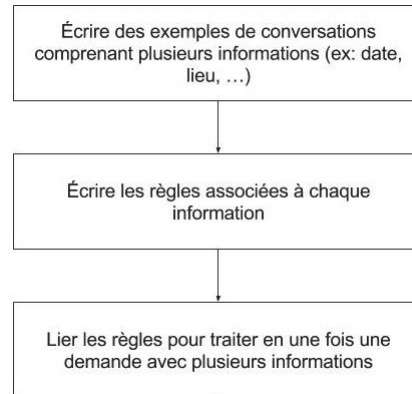


figure 2.5 – Étapes pour diversifier les interactions vocales pour simplifier les choix d’actions

2.4.2.1 Déterminer le réseau de neurones à appliquer

La revue de littérature du chapitre 1 a mis en avant les types de réseaux de neurones propres à la création de dialogues. Pour choisir le type de réseau de neurones à utiliser, il faut déterminer en parallèle si la mise en place manuelle d’un de ces réseaux est plus avantageuse que la réutilisation d’un réseau déjà conçu. La figure 2.6 reprend les étapes à suivre pour déterminer le réseau de neurones à appliquer.

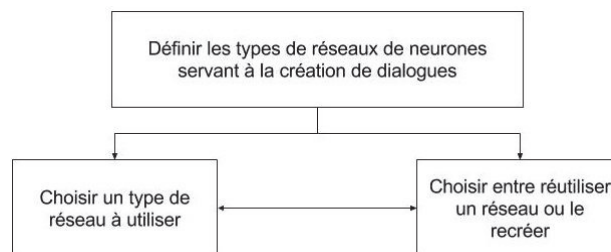


figure 2.6 – Étapes pour déterminer le réseau de neurones à utiliser

Avec le choix du réseau de neurones, certains paramètres sont inclus et vont permettre d’améliorer la qualité de l’apprentissage en les fixant à des valeurs. Le réseau de neurones choisi, il est ensuite nécessaire de fournir un corpus pour l’apprentissage.

2.4.2.2 Définir un corpus libre pour l'apprentissage du dialogue libre

Plus le corpus est grand, plus la capacité de comprendre et de répondre du réseau de neurones sera variée. Le corpus se choisit ou se construit selon différents critères :

- La langue ;
- La taille ;
- Le type de langage ;
- Les sujets souhaités.

La figure 2.8 montre les étapes à suivre pour choisir le corpus à adopter.

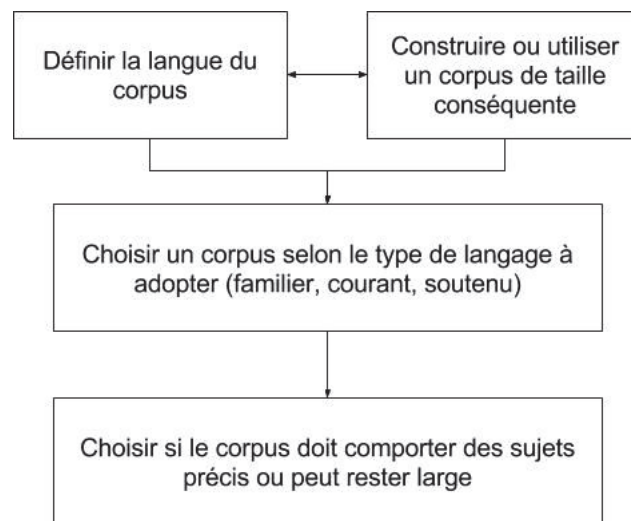


figure 2.7 – Étapes pour déterminer le corpus libre à utiliser

Le choix de la langue et le choix d'un corpus conséquent se font à un niveau égal car selon la langue désirée, il peut être difficile de trouver ou de créer un corpus assez grand. Bien que l'apprentissage ait pour but de mettre en place un dialogue libre, il est possible de choisir des sujets qui seront plus développés au sein du corpus, ce qui permettra au réseau de neurones d'avoir une discussion plus élaborée sur ces sujets. Selon la langue du corpus adoptée, il peut être nécessaire d'adapter la langue du corpus à celle des utilisateurs visés.

2.4. MÉTHODOLOGIE

2.4.2.3 Adapter la langue de l'apprentissage du réseau de neurones pour communiquer avec l'utilisateur

L'apprentissage réalisé, si la langue du corpus ne correspond pas à celle des utilisateurs, les réponses du réseau de neurones doivent être traduites. En réalité, le choix de traduire les réponses du réseau de neurones se fait en parallèle avec le choix de la langue du corpus. Par exemple, si un corpus anglophone est beaucoup plus large qu'un corpus francophone et que la langue visée est le français, il faut se demander si la traduction des réponses de l'anglais vers le français est suffisamment rapide et de bonne qualité. Le choix de la traduction se fait donc en plusieurs étapes, que l'on retrouve dans la figure 2.8.

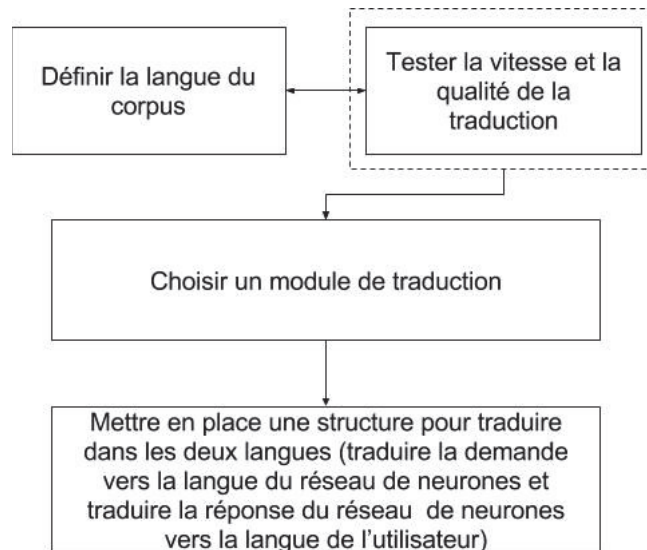


figure 2.8 – Étapes pour mettre en place la traduction et déterminer si c'est une solution viable

Lorsque ces étapes sont achevées, le dialogue libre peut être déployé au sein de l'application Amelis.

Le troisième objectif soulève le besoin d'intégrer deux modes d'interactions qui sont le vocal et le tactile. L'hypothèse d'une synchronisation simplifie leur cohabitation.

2.4.3 Synchroniser le vocal et le tactile

La synchronisation consiste en une communication entre les deux modes d'interaction pour qu'ils partagent les mêmes informations et que l'un puisse prendre le relais de l'autre. La synchronisation a lieu dans les deux sens mais d'une manière différente.

2.4.3.1 Synchroniser les liens du vocal au tactile

La synchronisation du vocal au tactile revient à envoyer les informations reçues par le vocal au tactile. Puisque seul le dialogue par règles permet à l'agent virtuel de contrôler le calendrier, les informations reçues pour le dialogue libre n'ont pas besoin d'être envoyées au tactile. De plus, certaines règles comme la salutation ou la demande de météo ne nécessitent pas de synchronisation. La synchronisation du vocal au tactile doit suivre plusieurs étapes que l'on retrouve dans la figure 2.9.

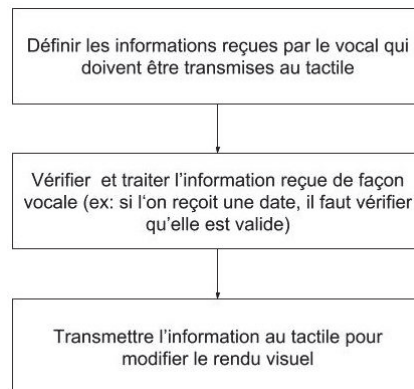


figure 2.9 – Étapes pour la synchronisation du vocal au tactile

La synchronisation en sens inverse suit un processus proche de celui-ci.

2.4.3.2 Synchroniser les liens du tactile au vocal

Comme précédemment, toute action tactile ne nécessite pas une synchronisation. Par exemple, la mise en marche manuelle de la musique ne peut être synchronisée avec le vocal, car le dialogue par règles ne gère pas les fonctionnalités de la musique. La figure 2.10 résume les étapes à suivre pour réaliser la synchronisation du tactile au vocal.

2.4. MÉTHODOLOGIE

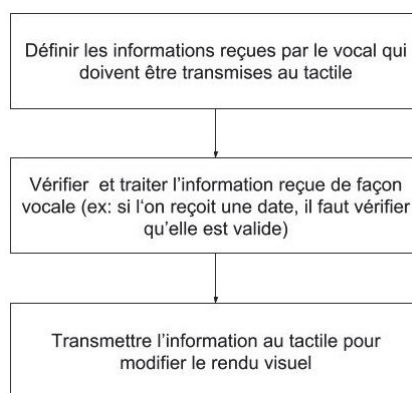


figure 2.10 – Étapes pour la synchronisation du tactile au vocal

Le dialogue par règles, le dialogue libre et la synchronisation doivent passer par plusieurs vérifications. On parle de tests techniques qui sont réalisés en même temps que la création des différentes fonctionnalités, mais aussi de validations auprès des utilisateurs.

2.4.4 Utiliser des tests pour vérifier sur un aspect technique le code et valider le fonctionnement du calendrier avec des utilisateurs

L'évaluation technique se faisant en parallèle avec la mise en place des fonctions, elle va être abordée avant les validations avec les utilisateurs.

2.4.4.1 Vérifier le dialogue par règles et la synchronisation par des tests "end-to-end"

Les tests "end-to-end" ou E2E permettent de simuler des appuis tactiles sur l'écran et de vérifier que les retours visuels sont bien présents. La méthode à adopter pour vérifier du code est de tester au fur et à mesure. Les tests E2E ont ici pour but de tester le dialogue par règles et la synchronisation entre le vocal et le tactile. Les tests réservés uniquement au tactile ne seront pas présentés ici. Tous les tests suivent un enchaînement similaire qui est résumé dans la figure 2.11.

Le dialogue libre ne peut être testé de la même manière.

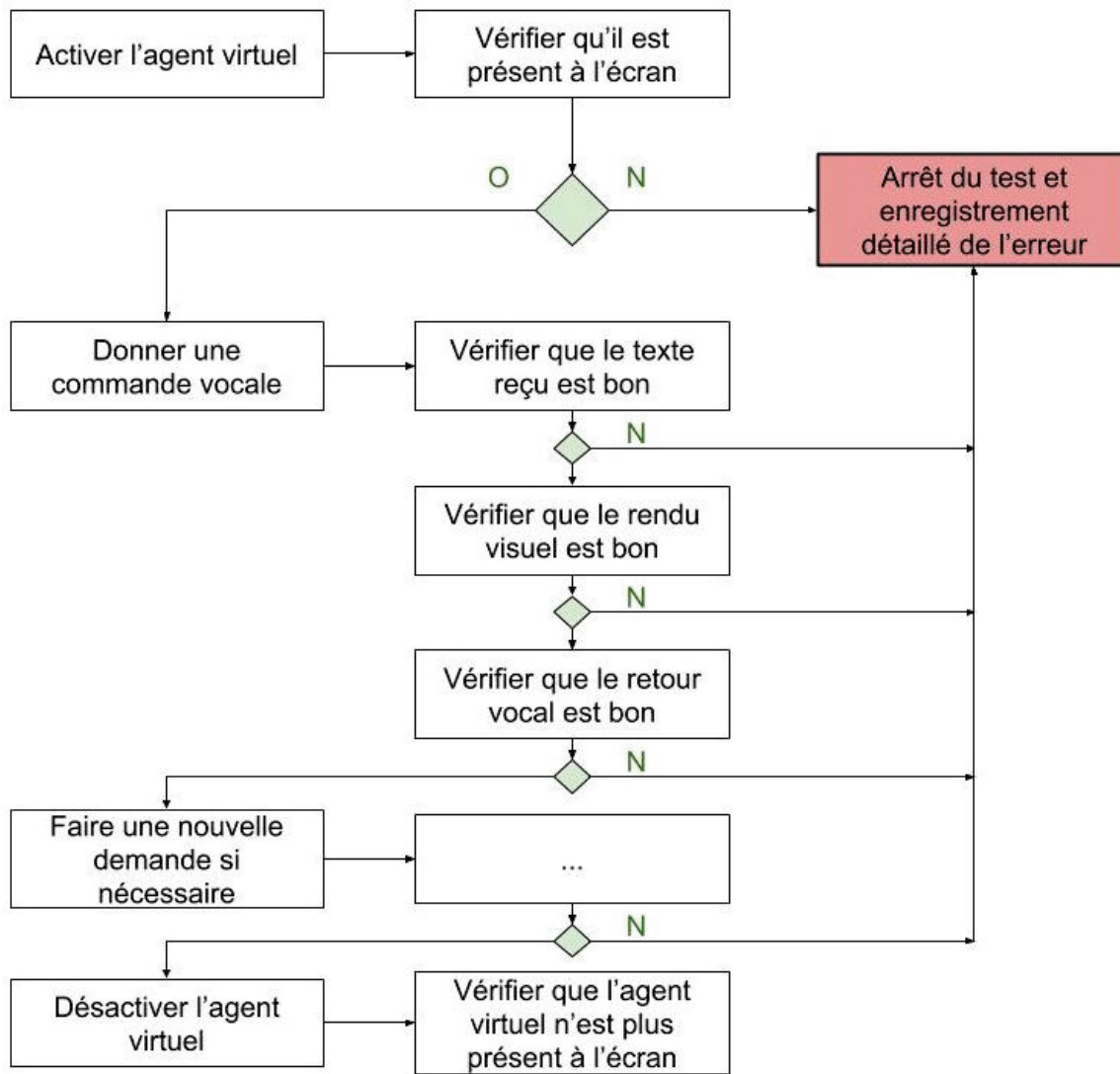


figure 2.11 – Étapes pour tester le dialogue par règles et la synchronisation avec des tests E2E

2.4. MÉTHODOLOGIE

2.4.4.2 Vérifier les apprentissages des réseaux de neurones en comparant leurs résultats

On parle ici d'un test de qualité d'un apprentissage et non d'un test des fonctions permettant de créer des réseaux de neurones. Les tests effectués consistent en une comparaison de plusieurs apprentissages de réseaux de neurones. Pour vérifier qu'un apprentissage est meilleur qu'un autre, on peut :

- Prendre un échantillon de phrases, par exemple 100 phrases, pour réaliser l'apprentissage automatique et comparer manuellement leur résultats ;
- Utiliser des tests BLEU (bilingual evaluation understudy) [90]. Les tests permettent d'évaluer automatiquement la qualité des réponses de l'apprentissage suivant le corpus de base. Dans l'article de Lavie [91], il est expliqué que le score BLEU se situe sur une échelle de 0 à 100. Pour un score entre 30 et 50, la qualité de réponse est considérée compréhensible et au dessus de 50, elle est considérée fluide et de bonne qualité ;
- Utiliser un analyseur de similarité de phrases en analysant la sémantique de deux phrases [92]. Cet analyseur prend en compte les informations sémantiques et l'ordre des mots pour calculer la similarité entre deux phrases.

Les tests techniques permettent de valider le bon fonctionnement du code et que l'application est fonctionnelle. La validation avec les utilisateurs permet quant à elle de savoir si les utilisateurs retrouvent ce qu'ils attendent et connaître leurs critiques par rapport à l'application Amelis.

2.4.4.3 Tester le niveau cognitif des utilisateurs

Avant de pouvoir valider le fonctionnement de l'application Amelis auprès des utilisateurs, un test, appelé "Montreal Cognitive Assessment" (MoCA) [93], est réalisé. Le test MoCA est conçu pour identifier le niveau cognitif de l'utilisateur. Le test évalue l'attention, la concentration, la mémoire, le langage, la visuoconstruction, la pensée conceptuelle, les calculs mentaux et l'orientation. Ce test dure une dizaine de minutes et est noté sur 30. Un utilisateur est considéré sans troubles cognitifs si son score est supérieur ou égal à 26. Pour les besoins de la première expérimentation, les participants sont sélectionnés s'ils atteignent un score au test MoCA supérieur à

20/30.

2.4.4.4 Tester le niveau de reconnaissance vocale des utilisateurs

En plus du test cognitif, un second critère de sélection des participants est la reconnaissance vocale. Les participants doivent lire six textes variant d'une trentaine à une soixantaine de mots. Trois doivent être lus sans pause (textes les plus courts) et les trois autres doivent être lus avec une pause à chaque fin de phrase. Le système de reconnaissance vocale reconnaît les six textes des participants. Le taux de reconnaissance vocale doit être supérieur à 70% de réussite pour qu'un participant soit sélectionné.

2.4.4.5 Valider le dialogue par règles auprès des utilisateurs

Le dialogue par règles est validé deux expérimentations auprès des usagers. Chaque protocole expérimental va ainsi être décrit.

Première expérimentation :

Cette expérimentation comprend une partie de trois rencontres avec les participants et une seconde partie où l'application Amelis est laissée à domicile chez la ou le participant(e) pour une durée de trois mois.

Lors des rencontres avec les participants, l'utilisateur doit réaliser des tâches tactiles et vocales pouvant être ou non sous la supervision du formateur. Lors des premières rencontres, le formateur montre comment réaliser une tâche, en tactile et en vocal, et demande ensuite à l'utilisateur de faire une tâche similaire.

Le formateur annonce à l'utilisateur s'il doit réaliser une tâche en tactile, en vocal ou en utilisant les deux. Lorsque le mode d'interaction n'est pas précisé par le formateur, l'utilisateur est libre de choisir celui qu'il préfère.

Durant chaque tâche, les observables sont les suivants :

- Le nombre et le type d'erreurs ;
- Les demandes d'assistance (partielle ou complète) par les participants ;
- Les réactions verbales et non verbales ;
- Les dysfonctionnements du système

2.4. MÉTHODOLOGIE

Une fois ces rencontres terminées, l'application Amelis est laissée au domicile des participants. Ils sont alors libres d'utiliser l'application Amelis comme ils le désirent.

Deuxième expérimentation :

Cette expérimentation se déroule également en deux parties comprenant des rencontres avec les participants et une autre partie où l'application Amelis est laissée à domicile. Lors de cette expérimentation, une moitié d'utilisateurs avait déjà utilisés Amelis et l'autre non. Lors des rencontres, le formateur explique et montre de manière brève le fonctionnement de l'application Amelis pour étudier la différence d'adaptation entre un utilisateur nouveau et un utilisateur habitué. Il est demandé aux participants de réaliser des tâches tactiles, vocales et des tâches mélangeant les deux modes d'interactions. En plus des observables de la première expérimentation, chaque tâche est chronométrée et l'utilisateur donne un retour en notant son ressenti suite à la réalisation de la tâche.

Pendant que l'application Amelis est à domicile, l'utilisateur est libre de l'utiliser. Cependant, des rappels automatiques provenant d'Amelis demandant de réaliser certaines fonctionnalités, s'activent durant cette semaine à domicile. Le mode d'interaction n'est pas précisé et peu de détails sont donnés à l'utilisateur. Par exemple : "Ajouter un rendez-vous à 16h".

2.4.4.6 Valider le dialogue libre auprès des utilisateurs

La validation du dialogue libre s'est faite lors de la deuxième expérimentation. Le protocole expérimental va être décrit ci-dessous.

Lors des rencontres avec les participants, il leur est demandé de discuter d'un sujet quelconque avec l'agent virtuel. Il est observé :

- La durée de l'échange ;
- Les réactions verbales et non verbales ;
- Les dysfonctionnements du système.

Lors de la semaine où l'application Amelis est laissée à domicile, des rappels demandent à l'utilisateur de parler d'un sujet libre avec l'agent virtuel.

Chapitre 3

Conception

Les objectifs et la méthodologie ont permis d'introduire le but de cette recherche et la manière dont elle a été mise en place. Ce chapitre sur la conception permet de développer et d'expliquer la réalisation permettant de répondre aux objectifs. Le plan de ce chapitre suit le plan de la méthodologie du chapitre [2](#).

3.1 Conception d'une architecture adaptable pour le dialogue et l'agent virtuel au sein de l'architecture d'Amelis

La construction de l'architecture adaptable intègre un dialogue et un agent virtuel à l'architecture d'Amelis. Pour expliquer l'architecture liée au vocal, il faut expliquer l'architecture de base d'Amelis liée au tactile.

3.1.1 Architecture tactile d'Amelis

Le calendrier Amelis est intégré à une tablette tactile. Toutes les fonctionnalités sont commandées par le tactile. L'architecture du tactile est décomposé par "pages" du calendrier Amelis (calendrier, photos, musique, fêtes, contacts, paramètres). Pour plus de clarté, nous allons définir le terme "application Amelis" qui servira à parler

3.1. CONCEPTION D'UNE ARCHITECTURE ADAPTABLE POUR LE DIALOGUE ET L'AGENT VIRTUEL AU SEIN DE L'ARCHITECTURE D'AMELIS

du calendrier Amelis, et le terme calendrier ou agenda qui sera utilisé pour parler de la page contenant le calendrier. Les fonctionnalités des pages sont expliquées dans le chapitre précédent. La figure 3.1 montre l'architecture des fonctionnalités tactiles.

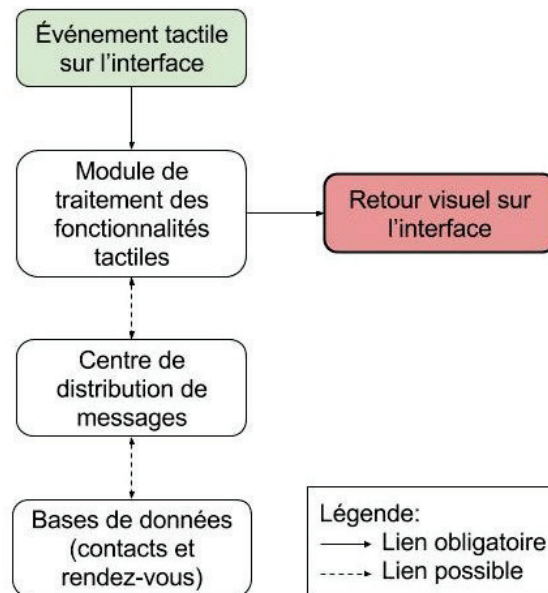


figure 3.1 – Architecture du tactile

Un événement tactile ne produit une réaction que lorsqu'une fonctionnalité est associée à un composant visuel. Par exemple, l'appui du bouton musique provoque l'apparition de la page des musiques, mais l'appui de la date se trouvant en haut au centre de l'application ne provoque aucune réaction puisqu'aucune fonctionnalité n'y est liée. Toute fonctionnalité tactile entraîne un retour visuel pour l'utilisateur. L'architecture du vocal fonctionne sur un principe similaire.

3.1.2 Architecture vocale d'Amelis

Pour une application tactile, il est simple, d'associer une fonctionnalité à un composant tactile. Pour le vocal, le traitement est plus délicat. Il faut récupérer la commande vocale et y appliquer un pré-traitement. Cela veut dire qu'il faut extraire des mots qui pourront être liés à des composants de l'application Amelis. Par exemple, "Quelle est la météo?" est une demande liée au composant de la météo. Des règles

dans le dialogue sont instaurées pour traiter ces événements. La figure 3.2 montre l'architecture du dialogue par règles traitant les événements vocaux. Néanmoins, à

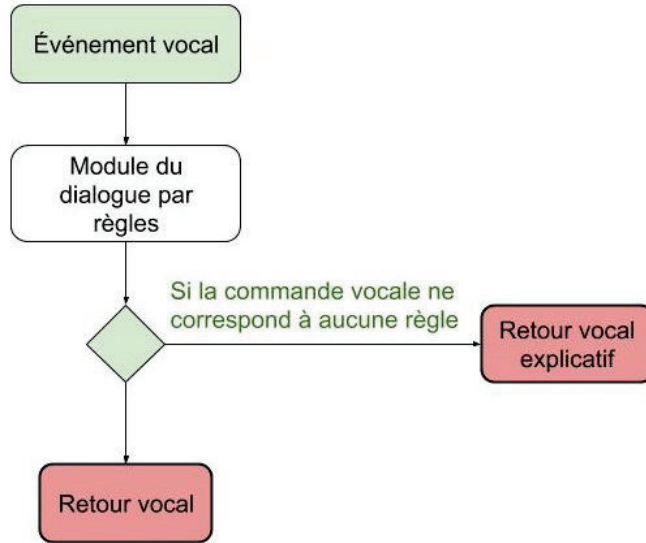


figure 3.2 – Architecture du vocal

l'inverse du tactile, les règles du vocal sont limitées à certains composants, et plus particulièrement ceux servant au traitement des rendez-vous. Reproduire toutes les actions tactiles en vocal avec un dialogue par règles est un travail large. C'est pourquoi dans cette recherche, les règles sont limitées à certaines fonctionnalités.

3.2 Conception du dialogue par règles

Le développement de la conception des règles suit la méthodologie exposé au chapitre 2.

3.2.1 Écriture des règles selon les fonctionnalités vocales de l'application Amelis

Les règles visent les fonctionnalités accessibles depuis la page du calendrier (voir figure 2.1 du chapitre 2 ou figure A.3 de l'annexe A). Ce choix a été fait pour limiter

3.2. CONCEPTION DU DIALOGUE PAR RÈGLES

le nombre d'interactions vocales et permettre un contrôle vocal centré sur les activités du calendrier.

3.2.1.1 Des règles centrées sur les rendez-vous

L'agent virtuel n'est présent que dans la page du calendrier. L'agent virtuel permet d'accéder à toutes les fonctionnalités de cette page. Le dialogue par règles permet donc d'interagir de façon vocale avec les menus de gauche, c'est-à-dire la navigation vers ces menus (voir figure 2.1 du chapitre 2 ou figure A.3 de l'annexe A), de connaître l'heure et la météo, et de saluer l'agent virtuel. Les autres règles sont dédiées au calendrier, ce qui comprend la navigation entre les mois et les années, et le contrôle des activités, c'est à dire ajout, modification, consultation et suppression de rendez-vous. Initialement, les règles ressemblaient à : une demande = une règle.

3.2.1.2 Construction initiale des règles

Pour construire des règles, le langage AIML, permettant de gérer la connaissance de bots ou d'agent virtuels, est utilisé. Ce langage est choisi, car il regroupe une grande communauté et évolue pour permettre de construire un dialogue à la fois simple et développé [94]. Un bot, nommé Mitsuku et construit à partir du langage AIML, a plusieurs fois remporté le prix Loebner qui est une compétition sur l'intelligence artificielle [94]. Les règles fonctionnent par un principe de reconnaissance de mots ou de modèles. Par exemple, une règle peut réagir au mot "météo", donc la phrase "Je veux voir la météo" active la règle s'activant lorsque le mot "météo est présent". Il est expliqué dans la méthodologie que des exemples de conversation ont été préalablement écrits. Cela a permis de trouver différentes formulations similaires pour une même tâche. Par exemple, pour les quatre formulations suivantes : "Je veux ajouter un rendez-vous" et "Peux-tu ajouter un rendez-vous?", ou "Je veux mettre une activité" et "Peux-tu mettre un rendez-vous", des modèles pour les règles sont définis. Il est constaté que "ajouter" et "rendez-vous" se répètent, de même que "mettre" et "activité". Pour ces exemples, deux règles ont été créées, une s'applique lorsque la demande de l'utilisateur comprend les termes "ajouter" et "rendez-vous", et une autre lorsque la demande de l'utilisateur comprend les termes "mettre" et "activité". L'inconvénient

est que chaque nouveau cas nécessite une règle. Les deux cas précédents peuvent être représentés de la manière suivante : R1aj (ajouter, rendez-vous) et R2aj (mettre, activité)

3.2.1.3 Réduire le nombre de règles

Le fait de réécrire une règle pour chaque nouveau cas multiplie le nombre de règles à écrire, alors que par exemple "ajouter rendez-vous" et "mettre activité" ont une même signification. Pour cela, AIML offre la possibilité d'utiliser des fichiers de synonymes. Au lieu qu'une règle s'active pour un mot, elle peut s'activer pour plusieurs.

Auparavant, les termes "ajouter" et "rendez-vous" activaient la règle "ajouter un rendez-vous". Maintenant, à l'aide de fichiers de synonymes, cette règle s'active désormais selon les mots se trouvant dans le fichier des synonymes à "ajouter" et le fichier des synonymes à "rendez-vous". La règle obtenue ressemble ainsi à : R1aj (ajouter, mettre), (rendez-vous, activité)

3.2.2 Mise en place des synonymes

Comme expliqué précédemment, AIML offre la possibilité de créer des fichiers de synonymes. Le nombre de règles est donc réduit, car chaque nouvelle formulation de la part de l'utilisateur ne nécessite plus forcément de rajouter une règle. Il suffit d'ajouter un synonyme à un des fichiers. Cependant, là encore se pose un problème car le nombre de synonymes et de formulations est grand. Il faut donc ajouter manuellement chaque nouvelle exception. Le dialogue par règles bloque à toute nouvelle exception tant qu'elle n'est pas traitée. Pour remédier à cela, un module a été conçu pour traiter automatiquement des synonymes courants.

3.2.2.1 Utilisation d'un dictionnaire externe

Les fichiers de synonymes d'AIML fonctionnent comme un dictionnaire. Dans un dictionnaire, à un mot sont associés plusieurs synonymes. Le module mis en place utilise un dictionnaire et traite de manière automatique les synonymes. Lorsqu'une demande de l'utilisateur ne correspond à aucune règle, cette demande est envoyée à ce module qui cherche un synonyme aux termes de la demande pour pouvoir activer une

3.2. CONCEPTION DU DIALOGUE PAR RÈGLES

règle. Par exemple, si une règle d'ajout répond au terme "ajouter" et la demande de l'utilisateur est "prendre un rendez-vous", cette demande est envoyée au dictionnaire qui lance une recherche est lancée pour savoir si un des termes de la phrase a pour synonyme "ajouter". Ici, le verbe "prendre" a pour synonyme "ajouter", cela permet de transformer la demande de l'utilisateur qui était "prendre un rendez-vous" pour "ajouter un rendez-vous". Cette nouvelle demande va cette fois-ci activer une règle. La figure 3.3 permet de voir l'évolution de l'architecture du vocal au sein d'Amelis.

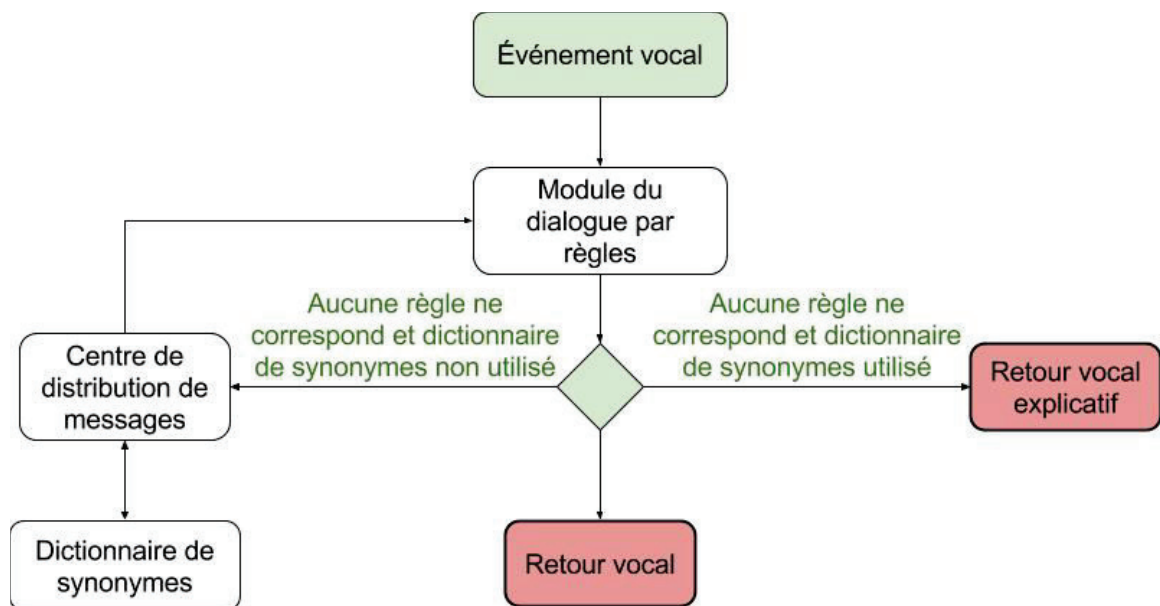


figure 3.3 – Architecture du dialogue par règles avec le traitement des synonymes

Les synonymes sont donc traités automatiquement et il n'y a plus besoin de les rentrer manuellement. Les fichiers de synonymes d'AIML sont maintenant destinés au traitement de cas particuliers comme les dates.

3.2.3 Traitement des dates et des heures par le dialogue par règles

Les formulations de dates ou d'heures sont très variées, si l'on prend en compte les formulations de l'utilisateur et la façon dont la reconnaissance vocale retranscrit la demande de l'utilisateur au système. Par exemple, "dix heures", "10h", "10 h", "10 :00"

correspondent à la même information. Les dates et des heures sont complexes à traiter avec les règles. L'utilisateur précise la date et l'heure d'un rendez-vous de façon très variée. Pour la date, les précisions oscillent entre un jour, un mois ou un jour, un mois et une année ou un mois ou un mois une année ou une année. De même, il peut être nécessaire de faire préciser s'il s'agit du matin ou de l'après-midi et ainsi que d'accepter une heure avec ou sans précision des minutes.

3.2.3.1 Traitement des dates

La gestion des dates nécessite l'usage d'un fichier de synonyme dans AIML, comprenant toutes les formulations possibles de dates. Par exemple, "demain", "5 janvier", "2 août 2015", Plusieurs règles ont ensuite été mises en place. Comme expliqué précédemment les différentes formes sous lesquelles une date peut être retrouvée, obligeant à instaurer des règles différentes. Par exemple, si le jour actuel est le 5 janvier et l'utilisateur dit "Je veux ajouter un rendez-vous le 10", cela doit être traité différemment que s'il avait dit "Je veux ajouter un rendez-vous le 10 février". En effet, dans le premier cas, il faut supposer que "le 10" correspond au 10 prochain ce qui veut dire le 10 janvier. Il en faut donc écrire une règle pour chaque cas dans la liste suivante :

- Un jour, ce qui implique de supposer que c'est le prochain jour à venir. Par exemple, "mardi" correspond à "mardi prochain", le "10" correspond au "10 prochain" ;
- Un jour et un mois, là encore, c'est le prochain "jour" du "mois" donné ;
- Un jour, un mois et une année ;
- Un mois, c'est à dire le prochain mois si ce n'est pas déjà le mois actuel ;
- Un mois et une année ;
- Une année.

Le traitement des heures, ne nécessite pas d'utiliser plusieurs règles. Que la demande comprenne simplement l'heure, l'heure et les minutes, ou juste les minutes, n'importe pas sur le traitement de la date. Cependant, certains cas nécessitent une précision. Il est courant de dire "il est onze heures" mais sans préciser si c'est le matin ou le soir. Une règle doit donc être mise en place pour demander à l'utilisateur s'il fait référence au matin ou à l'après-midi lorsqu'il donne une heure entre cinq

3.2. CONCEPTION DU DIALOGUE PAR RÈGLES

heures et onze heures. Les cas entre une heure et quatre heures ne sont pas pris en compte et sont automatiquement convertis en heures de l'après-midi, car les formulations d'heures sont propre aux rendez-vous et il est supposé que les utilisateurs ne mettront pas de rendez-vous entre une heure et quatre heures du matin.

Il n'est ici question que des règles. Elles permettent simplement de récupérer l'information, ce qui pose rapidement un problème. Aucune vérification n'est faite sur la date ou l'heure, à part le fait qu'elles sont reconnues par le dialogue par règles.

3.2.3.2 Vérification des dates et des heures

Il est nécessaire de vérifier que la date reçue a un format valide. Cependant, plusieurs vérifications et ajustements sont nécessaires. Il est dit dans la liste précédente qu'il était possible de donner un jour ou un jour et un mois,

Le premier ajustement consiste à définir la date avec le jour, le mois et l'année pour permettre les traitements des rendez-vous. Si seul le jour est reçu, on suppose que c'est le prochain jour de ce mois de cette année. Un fonctionnement similaire s'applique pour toute date qui ne contient pas le jour, le mois et l'année.

Un autre point est la validation de l'existence de la date. Par exemple, la date jeudi 16 mars 2018 n'existe pas. La date n'est pas valide à un jour près (jeudi 15 mars 2018). Il est alors possible de soit faire un retour vocal expliquant à l'utilisateur que la date n'existe pas, puis de lui suggérer qu'il voulait peut-être parler du jeudi 15 mars 2018, soit de ne prendre en compte comme date que le 16 mars 2018, en supposant que l'utilisateur s'est trompé sur le jour "jeudi" mais que la date est bonne. La deuxième solution est pour l'instant mise en place car la première nécessite d'avoir des règles comprenant le jour nommé, le jour chiffré, le mois et l'année, mais il n'est pas précisé à l'utilisateur que la date est mauvaise puisque le jour est ici omis. Il faut également vérifier les dates voisines de celle reçue pour faire des propositions. La deuxième solution est donc plus rapidement et facilement utilisable. De plus, l'utilisateur peut modifier la date s'il le souhaite, mais cela ne lui est pas dit explicitement, l'utilisateur doit penser à demander à modifier la date.

Une dernière vérification au niveau des dates et des heures, est nécessaire lors de la manipulation des rendez-vous. Si l'utilisateur donne une date valide, il faut regarder, dans le cas d'un ajout ou d'une modification d'activité, deux choses :

- Que la date ne se situe pas dans le passé. Par exemple, le 7 janvier à 15h00, il n'est pas possible d'ajouter un rendez-vous dans le passé. Donc la date et l'heure fournies par l'utilisateur doivent se trouver dans le futur ;
- Qu'il n'y a pas de conflits horaires entre les rendez-vous. Il faut vérifier, qu'en ajoutant ou en déplaçant un rendez-vous, il n'y ait pas une superposition ;

Pour la vérification de la date dans le passé, il suffit de comparer la date et l'heure actuelles avec la date et l'heure fournies. Si la date se trouve dans le passé, l'agent virtuel demande à l'utilisateur de reformuler la date, et si la date est la même que le jour actuel mais que l'heure est dans le passé, l'agent virtuel demande une nouvelle heure.

Pour les conflits d'horaires, il est décidé que les rendez-vous ont une durée d'une heure. La complexité d'interaction augmente si l'utilisateur doit choisir une heure de début et une heure de fin pour le rendez-vous. Or le public visé des aînés n'est pas toujours habitué à la technologie. Lors de la réception d'une date validée, il y a une vérification pour savoir s'il y a d'autres rendez-vous dans la même journée. Si c'est le cas, il est vérifié que l'heure reçue a au minimum une heure d'écart avec le ou les rendez-vous existants. Ainsi si un rendez-vous est inscrit à 16h, l'utilisateur ne peut pas ajouter un nouveau rendez-vous entre 15h01 et 16h59 pour un nouveau rendez-vous. Dans le cas d'un conflit, l'agent virtuel annonce à l'utilisateur l'heure du rendez-vous qui pose conflit et demande à ce que l'heure soit reformulée. La figure 3.4 résume les modules de traitement des dates et la figure 3.5, l'évolution de l'architecture vocale.

L'utilisateur peut dialoguer avec l'agent virtuel en utilisant les formulations qu'il désire. Néanmoins, il est encore contraint de suivre un schéma de dialogue défini par les règles.

3.2.4 Augmentation des choix d'actions

Bien que l'utilisateur soit libre de la formulation de sa demande, cette dernière ne sera pas toujours traitée entièrement à cause d'un traitement au cas par cas avec les règles. Par exemple, l'ajout de rendez-vous comporte plusieurs champs : date, heure, titre, catégorie, Chaque règle est conçue pour récupérer un champ. Une

3.2. CONCEPTION DU DIALOGUE PAR RÈGLES

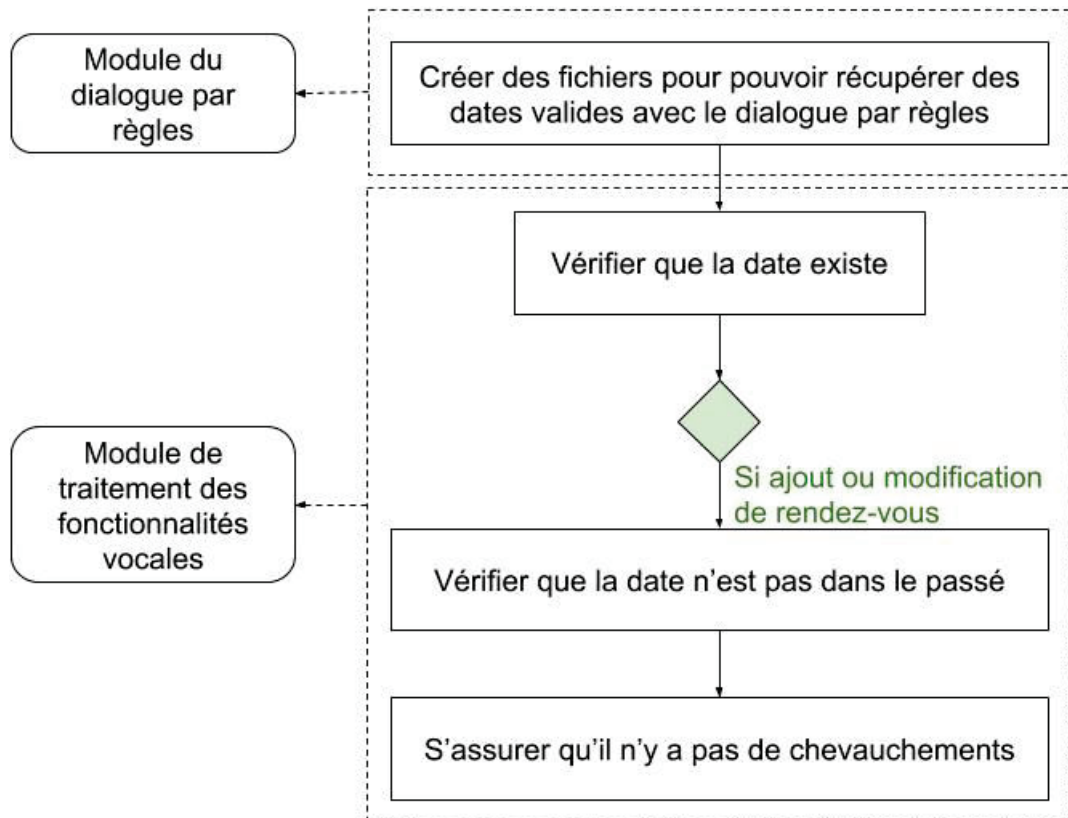


figure 3.4 – Étapes pour traiter les dates avec le dialogue par règles

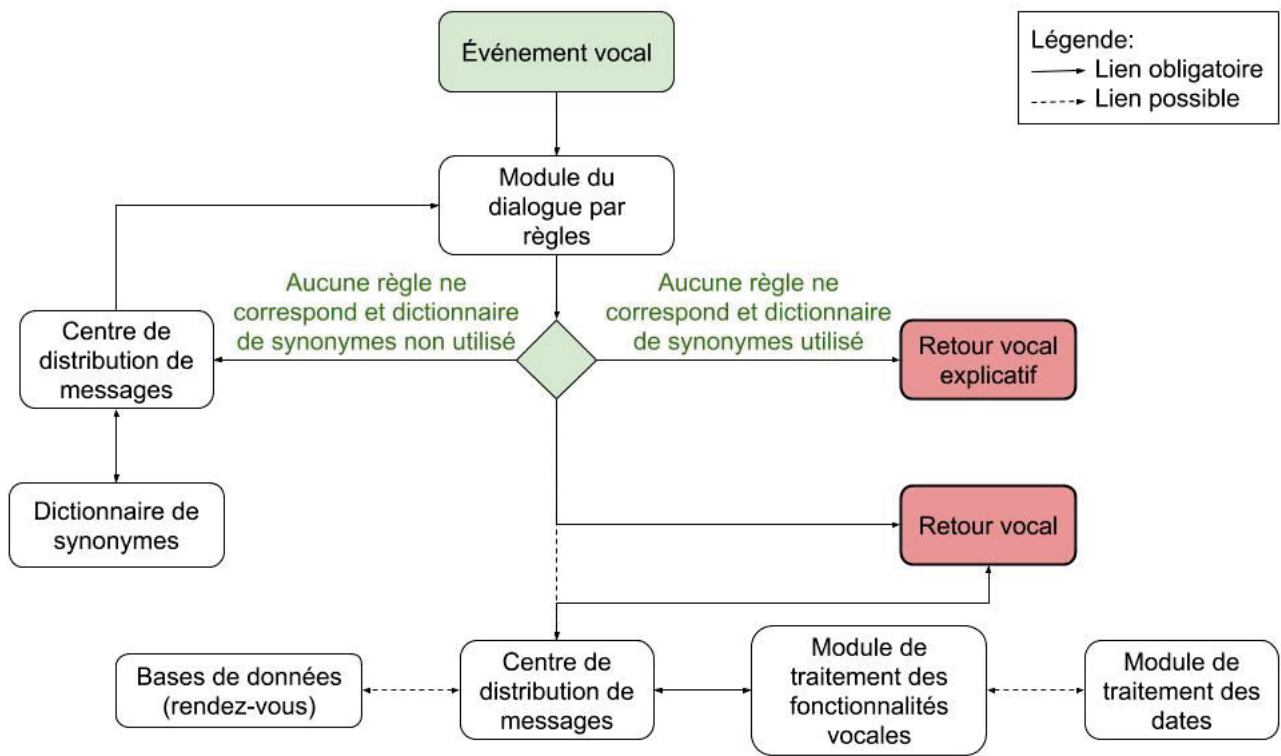


figure 3.5 – Architecture du dialogue par règles avec le traitement des dates

3.2. CONCEPTION DU DIALOGUE PAR RÈGLES

une règle peut être conçue pour traiter plusieurs champs à la fois mais cela multiplie les possibilités. Par exemple, une règle qui gère juste les dates, une autre qui gère les dates et les heures, et ainsi de suite. En gardant des règles adaptées pour un champ, une multiplication du nombre de règles est évitée, mais cela impose d'utiliser un ordre pour récupérer les informations. L'utilisateur doit donner la date, une fois validée, il doit donner l'heure, et le processus se répète pour les autres champs. L'utilisateur ne peut donc pas formuler une demande comme "Je veux ajouter un rendez-vous famille demain à 18h", car seule la date est traitée en premier, puis l'agent virtuel redemande les informations pour les champs suivants.

3.2.4.1 Refactorisation des règles pour élargir les choix

Pour éviter une multiplication du nombre de règles tout en offrant à l'utilisateur la possibilité de formuler la demande de son choix, le traitement des règles doit être revu. Pour cela, la structure des règles est modifiée. Au lieu de traiter une règle à la fois, une imbrication de règles permet d'en traiter plusieurs à la fois. Le fonctionnement des règles avant et après l'imbrication des règles est expliqué dans les deux figures 3.6 et 3.7, qui représentent le début d'un ajout vocal d'un rendez-vous.

La structure avec imbrication des règles permet d'envoyer la demande de l'utilisateur dans une boucle de vérification pour récupérer plusieurs informations à la fois. Cependant, cette boucle n'est pas applicable pour chaque type d'information. Par exemple, le dialogue par règles permet de reconnaître quand l'utilisateur parle d'une date lorsque des fichiers sont créés pour faire ce traitement, mais il est impossible ou très difficile de connaître un champ tel que le titre. Si l'utilisateur dit "Ajouter rendez-vous tournoi de pétanque demain". Les règles ne permettent pas de détecter que "tournoi de pétanque" puisse être un titre. Il faudrait pour cela étudier les habitudes des utilisateurs et trouver les titres qui reviennent fréquemment.

Une autre refactorisation effectuée est l'augmentation des fonctionnalités couvertes par le dialogue par règles. L'agent virtuel n'étant présent que dans la page du calendrier, il permet d'accéder aux autres pages de l'application Amelis. Cette refactorisation concerne l'élargissement des possibilités d'actions avec ses pages et la mise en place d'une aide vocale plus importante.

Comme les fonctionnalités accessibles en vocal ne concernent que la page du ca-

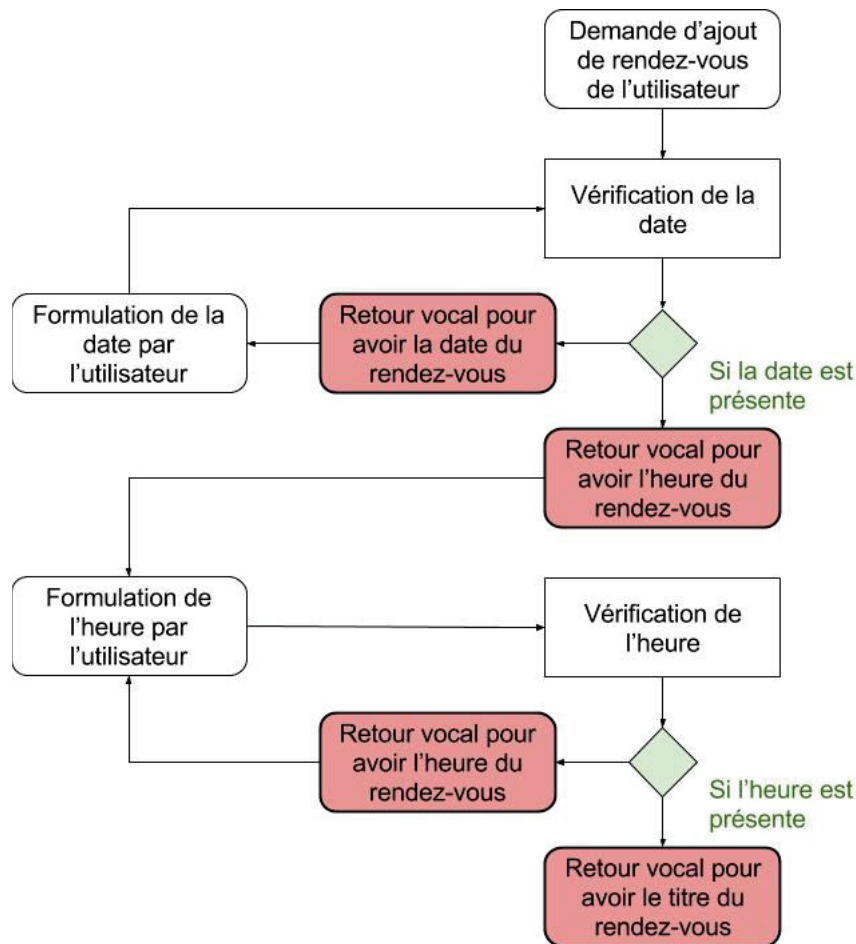


figure 3.6 – Structure des règles avec un traitement au cas par cas

3.2. CONCEPTION DU DIALOGUE PAR RÈGLES

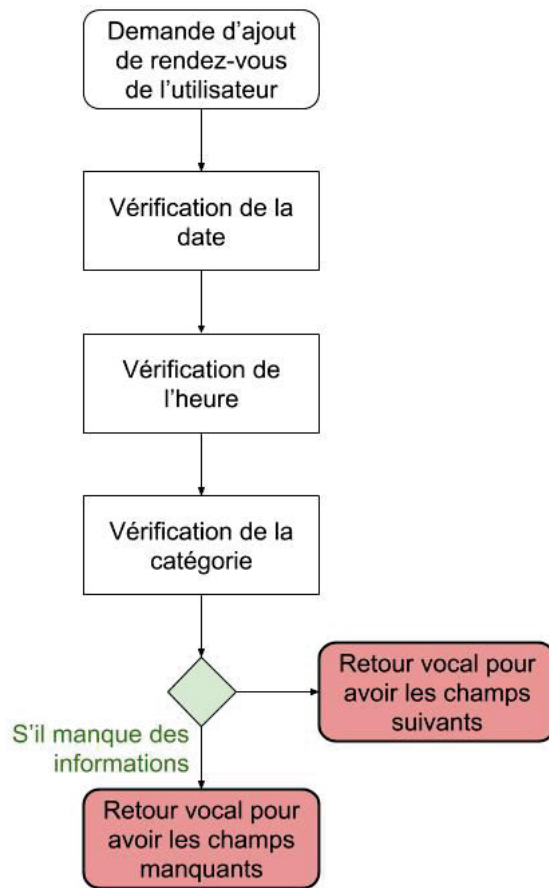


figure 3.7 – Structure des règles avec une imbrication des règles

lendrier, aucune règle ne permet de réaliser une fonctionnalité d'une autre page. Il est simplement possible de demander d'aller dans une autre page. Les règles ne sont donc conçues que pour comprendre une demande de changement de page. Si l'utilisateur dit "je veux voir mes photos", la page des photos s'affiche, mais si la commande est "Mets en marche la musique", la fonctionnalité d'activation vocale de la musique n'étant pas mise en place, l'agent virtuel annonce qu'il ne comprend pas la demande.

Pour gérer ce problème, il faut soit construire des règles permettant d'effectuer plus d'actions autres que celles spécifiées, soit rester dans un cadre de règles définies. Pour garder une simplicité au niveau du calendrier pour les aînés, le cadre des règles restreignant les actions avec la page du calendrier doit rester en place. Néanmoins, il est possible d'élargir la compréhension des demandes pour ces pages. Cela veut dire que même si la demande ne correspond pas à une demande d'accès à la page, par exemple "jouer de la musique", la demande doit être traitée en rapport avec la page souhaitée. En plus des règles permettant de naviguer entre les pages, des règles sont mises en place pour comprendre si une demande un rapport avec une des pages de l'application Amelis et de renvoyer l'utilisateur vers cette page en lui expliquant que l'action qu'il désirait ne peut être effectuée qu'en tactile. Prenons ces deux exemples :

Exemple1 :

Utilisateur (U) : Je veux voir ma musique.

Agent Virtuel (AV) : Je vous affiche la musique.

Exemple 2 :

U : Je veux écouter de la musique.

AV : Cette action n'est possible qu'en tactile mais je vous renvoie vers la page de la musique.

Comme toutes les fonctionnalités du calendrier ne sont pas accessibles en vocal, de nouvelles règles permettent d'expliquer pourquoi cela n'est pas accessible et renvoie l'utilisateur vers la page où ces fonctionnalités se trouvent.

Un problème reste néanmoins présent. Le dialogue par règles se concentre sur les fonctionnalités de l'application Amelis, mais ne gère pas les cas en dehors des fonctionnalités. Si l'utilisateur demande à l'agent virtuel s'il aime regarder des films, ce dernier ne peut comprendre puisque cela ne se trouve dans aucune règle.

3.2. CONCEPTION DU DIALOGUE PAR RÈGLES

3.2.5 Impasse du dialogue par règles sur les sujets libres

Pour pouvoir répondre à des sujets libres qui n'appartiennent pas au dialogue par règles, il faut une solution qui apprenne de l'utilisateur ou une solution qui permette déjà de dialoguer de manière libre.

3.2.5.1 Apprentissage automatique à partir des règles

Une première solution est l'utilisation d'un apprentissage au niveau des règles. Pour chaque entrée n'étant pas gérée par le dialogue par règles, l'agent virtuel propose à l'utilisateur d'apprendre une réponse à sa demande. Par exemple, l'utilisateur dit "J'aime les bateaux", et l'agent virtuel ne peut pas répondre car cela n'est pas traité par les règles. La solution est d'expliquer à l'utilisateur que l'agent virtuel ne peut pas répondre à cette demande pour l'instant mais qu'il souhaite apprendre la réponse à donner. Par exemple, l'agent virtuel dit "Je ne comprends pas votre demande, qu'aurais-je dû répondre à 'J'aime les bateaux' ?" et l'utilisateur donne alors la réponse que l'agent virtuel doit retenir. Mais l'inconvénient est que pour une demande où une réponse y est associée, la même demande qui change d'un mot ou plus ne fonctionne pas. Pour éviter ça, un module sémantique est mis en place pour découper la phrase. Dans une phrase, tous les mots ne sont pas obligatoirement importants. Une règle est construite pour simplement reconnaître les mots importants, comme les verbes. Par exemple, la phrase "J'aime les bateaux", la règle va réagir au verbe "aimer" et au nom "bateau". La phrase "J'aimerais aller voir des bateaux" serait donc reconnue par la règle.

Cette solution n'est pourtant pas optimale pour plusieurs raisons :

- Pour chaque nouvelle entrée non reconnue, il faut demander à l'utilisateur une réponse ;
- Si la règle créée réagit à certains mots, il se peut qu'elle empiète sur une autre règle qui doit proposer une réponse différente ;
- Il est compliqué de savoir quand commence la réponse à enregistrer qui est donnée par l'utilisateur. Par exemple "Quelle doit être la réponse à 'J'aime les bateaux' ?", l'utilisateur répond "Je veux que la réponse soit : 'Quel type de bateau aimes-tu ?'". Il n'y a pas moyen de savoir que la réponse à enregistrer

est juste la deuxième partie de la phrase ;

- L'utilisateur peut se tromper dans sa réponse ou la reconnaissance vocale peut ne pas avoir reconnu correctement la réponse, il faut donc redemander la réponse ;

L'apprentissage par les règles n'est donc pas une solution viable sans suffisamment de vérifications. Un module de dialogue capable de parler préalablement de sujets libres serait préférable.

3.3 Création d'un dialogue libre à l'aide d'un réseau de neurones

Dans le chapitre 1 de la revue de la littérature, les différents réseaux de neurones propres à la construction d'un dialogue ont été détaillés. Il a été montré que les modèles d'attention sont plus efficaces pour bâtir un dialogue. Les modèles de réseaux de neurones sont variés mais il n'a pas été encore déterminé la supériorité d'un modèle par rapport à un autre. Deux modèles d'attention déjà existants ont été choisis.

3.3.1 Utilisation de deux réseaux de neurones libres d'accès

3.3.1.1 Modèle proposé par Google

Le modèle de Google, appelé tf-seq2seq [95], utilise les bibliothèques de TensorFlow développées par Google. TensorFlow est libre d'utilisation et regroupe une large communauté d'utilisateurs, comprenant des chercheurs et des professionnels industriels. De plus, la documentation disponible, ainsi que les nombreux tutoriels sur internet permettent de comprendre TensorFlow et les fonctions utilisées. Le modèle de Google comprend différents réseaux de neurones adaptés pour différentes tailles de corpus et est mis à jour régulièrement.

3.3.1.2 Modèle PyTorch

PyTorch est développé par Facebook. PyTorch est plus récent que TensorFlow et regroupe une plus petite communauté. Cependant, son efficacité et sa vitesse sont

3.3. CRÉATION D'UN DIALOGUE LIBRE À L'AIDE D'UN RÉSEAU DE NEURONES

adaptées pour de petits projets de recherche. PyTorch propose plusieurs modèles adaptés selon les besoins [96].

Plusieurs sites proposent une comparaison entre TensorFlow et PyTorch. Pour ce projet, les apprentissages sont réalisés à l'aide de ces deux solutions. Cela permet de comparer quel apprentissage est le plus adapté pour la mise en place d'un dialogue libre.

Les deux modèles d'apprentissage n'ont pas été modifiés pour le projet. Le corpus des apprentissages et les variations des paramètres des réseaux de neurones pour les rendre plus performants ont été adaptés au projet.

3.3.2 Choix d'un corpus

Les réseaux de neurones propres au dialogue utilisent un corpus de textes sous forme de questions réponses pour l'apprentissage. Plus ce corpus est important, plus le réseau de neurones aura des "connaissances variées".

3.3.2.1 Conception d'un corpus francophone

L'application Amelis vise un public québécois et français, le dialogue entre l'agent virtuel et l'utilisateur est donc en français. Pour trouver des textes sous forme de questions réponses, les scripts de films et de pièces de théâtre sont idéaux. Le corpus francophone est constitué de bribes de scripts de films variés, ainsi que de différents extraits de pièces de théâtre. Le corpus atteint 10000 phrases. Cette taille de corpus est insuffisante pour avoir un apprentissage large. La taille du corpus est réduite car des lois françaises imposent une restriction sur les droits d'accès aux scripts. Il est rare de trouver des scripts complets et cela limite donc la conception d'un corpus francophone.

A l'inverse, beaucoup de scripts anglophones sont disponibles et il est possible de constituer un corpus conséquent.

3.3.2.2 Utilisation d'un corpus anglophone

Dans les modèles de réseaux de neurones libres d'accès dédiés au dialogue, des corpus anglophones de grande taille sont parfois fournis. Il est donc beaucoup plus

simple d’avoir un grand nombre de données pour l’apprentissage. Ainsi, l’apprentissage sera réalisé en anglais et non en français. C’est pourquoi il faut utiliser un module de traduction.

3.3.2.2.1 Mise en place d’un traducteur Avant d’utiliser un corpus anglophone, il faut s’assurer de la viabilité de la solution et du temps de traitement des traductions pour qu’il ne ralentisse pas le système. Si l’agent virtuel met un temps trop important à répondre, on perd la convivialité du dialogue. La traduction doit être réalisée dans deux sens. Il faut traduire la phrase de l’utilisateur, du français vers l’anglais, pour l’envoyer au réseau de neurones et ensuite traduire la réponse du réseau de neurones, de l’anglais vers le français, pour l’envoyer à l’agent virtuel.

Le module de traduction utilisé se sert de l’API de Google translate. Cette API est utilisée car Google travaille depuis longtemps sur son module de traduction pour le faire évoluer et la qualité des traductions ne cesse de s’améliorer [97]. Pour vérifier la rapidité de l’API, deux corpus, un en français et l’autre en anglais, d’une vingtaine de phrases plus ou moins longues sont utilisés. La traduction du français vers l’anglais, puis de l’anglais au français prend en moyenne une seconde. La vitesse de traduction est donc suffisamment rapide.

Le module de traduction est efficace pour utiliser un corpus anglophone.

3.3.2.2.2 Choix du corpus anglophone Les scripts de films anglophones sont courants, cependant il reste un traitement à réaliser sur ceux-ci. Il faut avoir un corpus sous forme de questions réponses, on ne peut donc pas prendre un script et l’utiliser tel quel. Comme évoqué précédemment, il existe des corpus déjà traités qui sont fournis avec les modèles libres d’accès. Le corpus, utilisé pour les apprentissages, contient des conversations extraites d’oeuvres de fiction. Ce corpus comprend 220000 échanges entre différents interlocuteurs provenant de 600 films. Ce corpus est assez large pour être utilisé pour un apprentissage. De plus, les phrases de ce corpus sont exprimées en langage courant, ce qui correspond au langage recherché pour l’agent virtuel.

L’apprentissage pour le dialogue libre peut donc débuter. Plusieurs apprentissages vont être effectués en faisant varier les paramètres des réseaux de neurones pour

3.3. CRÉATION D'UN DIALOGUE LIBRE À L'AIDE D'UN RÉSEAU DE NEURONES

trouver la meilleure combinaison pour établir un réseau de neurones performant et optimisé.

3.3.3 Variation des paramètres

Dans la liste ci-dessous se retrouvent les différents paramètres modifiés pour optimiser l'apprentissage :

- L'encodeur et le décodeur. L'encodeur prend une phrase d'entrée qu'il encode une représentation de cette phrase. La représentation encodée est ensuite transmise au décodeur pour obtenir une réponse à la phrase d'entrée. La figure 3.8 schématise le fonctionnement ;
- Le nombre de couches. Chaque couche peut appliquer une fonction à la couche précédente pour produire une phrase de sortie. Les fonctions appliquées dépendent du réseau utilisé ;
- Le taux d'apprentissage. Il fait converger la fonction d'apprentissage pour obtenir le meilleur apprentissage. En effet, un minimum doit être atteint pour que la fonction d'apprentissage soit optimale. Dans la figure 3.9, trois utilisations du taux d'apprentissage sont présentées. Les flèches représentent un taux constant durant l'apprentissage. En pointillé, le taux d'apprentissage est grand pour essayer d'arriver au point de convergence en sautant de part en part de la fonction. Les arcs représentent un taux d'apprentissage élevé au départ pour atteindre plus rapidement le point de convergence, puis il diminue pour trouver le point de convergence de manière plus précise ;
- L'optimisateur. Il sert à minimiser (ou maximiser selon le but) une fonction objective qui est une fonction "Erreur". L'idée est de modifier les paramètres de cette fonction pour réduire la perte lors de l'apprentissage.

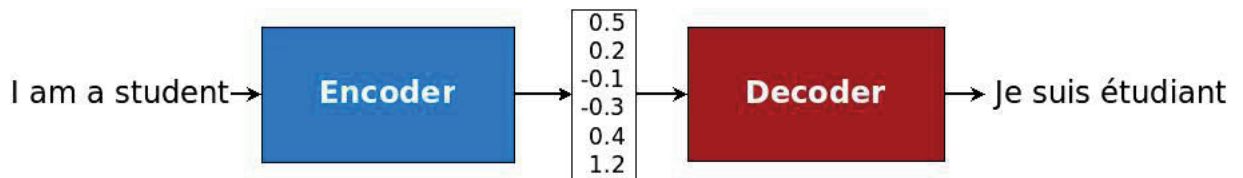


figure 3.8 – Architecture d'un encodeur décodeur

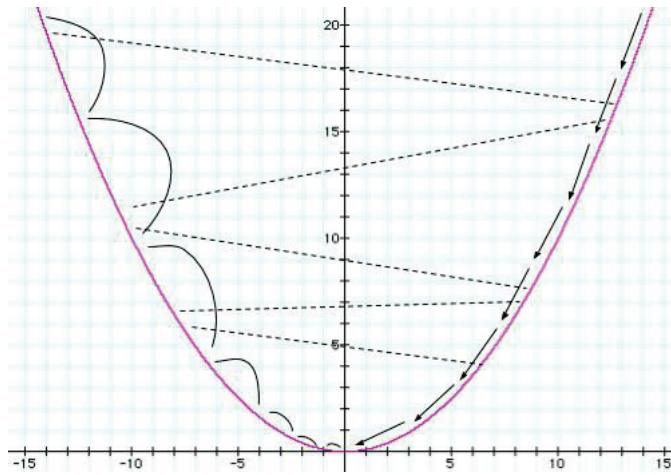


figure 3.9 – Convergence de la fonction d'apprentissage

L'application Amelis comporte maintenant un module de dialogue libre et un module de dialogue par règles pour contrôler l'interface de l'application Amelis. La figure 3.5 montrait l'architecture du dialogue par règles. Dans la figure 3.10, il est présenté l'architecture du dialogue libre. Lors de la réception d'un événement vocal, celui-ci est d'abord traité par le dialogue par règles pour détecter une éventuelle demande de l'utilisateur par rapport au calendrier. Si ce n'est pas le cas, la demande de l'utilisateur est traitée par le dialogue libre. La sortie "Retour vocal explicatif" de la figure 3.5 est alors remplacée par le dialogue libre.

Comme expliqué dans la méthodologie, l'interaction entre l'utilisateur et l'application est améliorée par la mise en place d'une synchronisation entre le tactile et le vocal.

3.4 Conception de la liaison entre le vocal et le tactile

La communication entre les deux modes d'interaction tactile et vocal doit avoir lieu dans les deux sens. Cependant, chaque sens a son propre mode de fonctionnement.

3.4. CONCEPTION DE LA LIAISON ENTRE LE VOCAL ET LE TACTILE

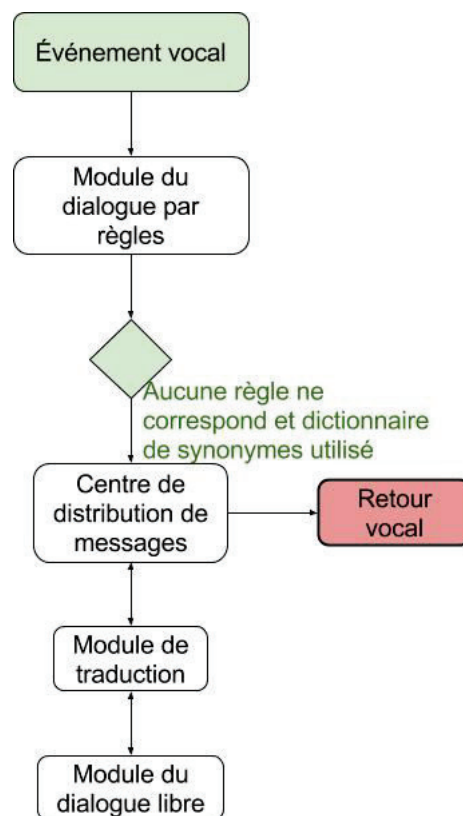


figure 3.10 – Architecture du dialogue libre

3.4.1 Synchronisation du vocal vers le tactile

La synchronisation du vocal vers le tactile active une fonctionnalité tactile, déjà mise en place et déjà fonctionnelle, à partir d'une action commandée par le vocal. Le dialogue par règles permet déjà d'extraire des informations. La figure 3.11 reprend l'exemple d'un début d'ajout de rendez-vous et montre comment sont transmises les informations vocales pour activer une fonctionnalité tactile et ainsi donner le retour visuel attendu.

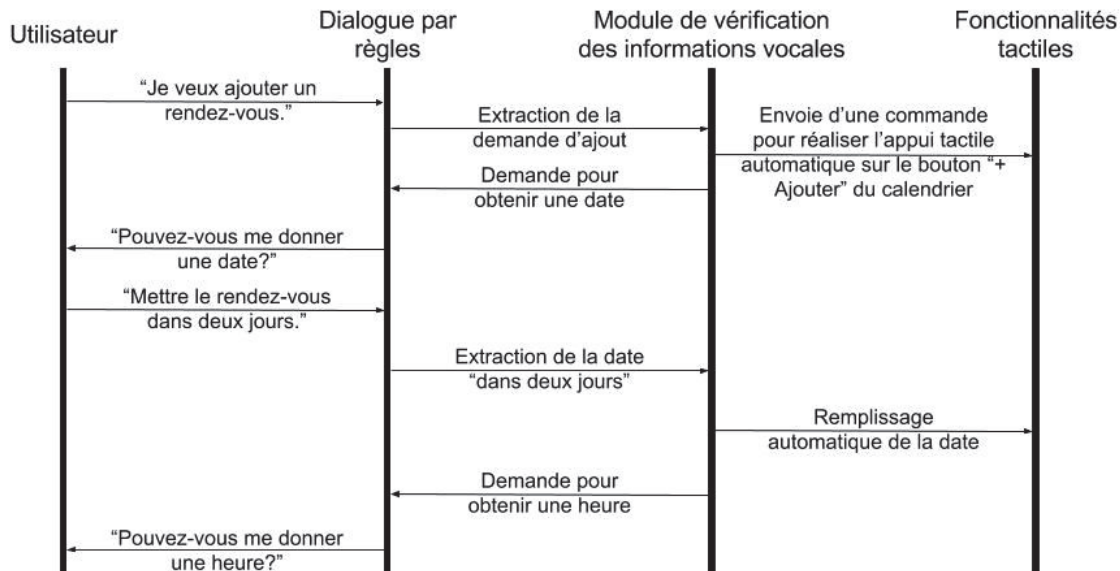


figure 3.11 – Exemple de synchronisation du vocal vers le tactile

Cette synchronisation n'est cependant utilisée que pour la page du calendrier, puisque l'agent virtuel n'est présent que dans celle-ci. Elle n'a donc lieu que pour le traitement d'un rendez-vous, pour la navigation dans le calendrier ou pour le changement de page.

La figure 3.12 reprend l'évolution de l'architecture du vocal en incluant la synchronisation du vocal vers le tactile.

Cette synchronisation ne se réalise que si les informations vocales sont valides. Par exemple, lors d'un ajout de rendez-vous, si la date fournie n'est pas valide, la date ne se synchronise pas. La synchronisation en sens inverse est similaire.

3.4. CONCEPTION DE LA LIAISON ENTRE LE VOCAL ET LE TACTILE

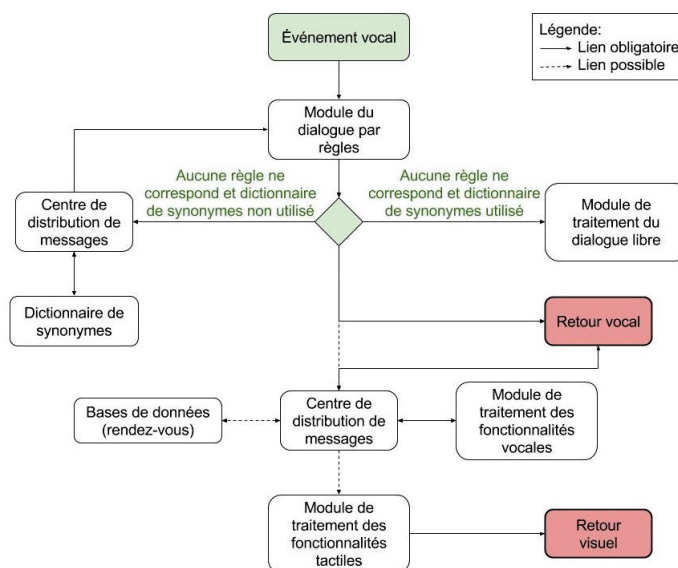


figure 3.12 – Architecture du vocal comprenant la synchronisation du vocal vers le tactile

3.4.2 Synchronisation du tactile vers le vocal

Pour chaque action tactile qui existe en vocal, des données sont envoyées pour effectuer une synchronisation. Comme pour la synchronisation du vocal vers le tactile, la validité des informations tactiles reçues est vérifiée. Pour que la synchronisation ait lieu, lorsqu'une commande tactile est récupérée, elle est transformée en phrase à envoyer au dialogue par règles. Par exemple, l'utilisateur clique sur le bouton "+ Ajouter" du calendrier, l'information "ajouter rendez-vous" est envoyée au dialogue par règles.

Une vérification supplémentaire prend place lors de l'envoi d'informations vers le vocal. Il est vérifié si l'agent virtuel est actif ou non. S'il l'est, il va alors pouvoir donner la réponse suite à l'activation d'une règle, mais s'il n'est pas actif, la règle est tout de même activée car l'information doit être fournie pour que les deux modes d'interactions soient rendus au même niveau. Par contre, l'agent virtuel ne donnera pas de réponse.

Enfin, dans le cas où un champ est complété en tactile, par exemple le titre, l'application Amelis attend que l'utilisateur finisse de rentrer son titre pour ensuite

l'envoyer au vocal. En effet si la synchronisation avait lieu à chaque nouveau caractère et que l'agent virtuel était activé, il donnerait une réponse à chaque nouveau caractère rentré.

La figure 3.13 reprend l'évolution de l'architecture du tactile en incluant la synchronisation du tactile vers le vocal.

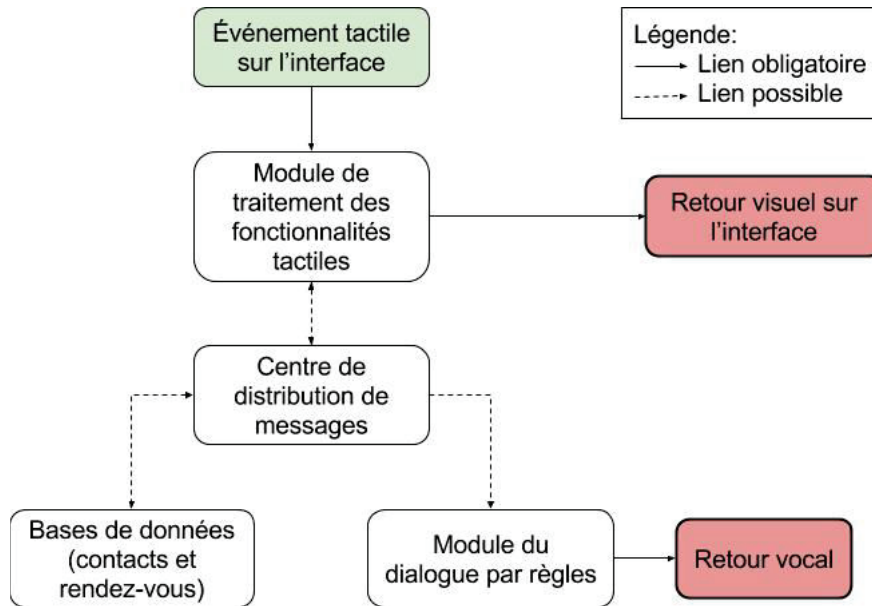


figure 3.13 – Architecture du tactile comprenant la synchronisation du tactile vers le vocal

Les fonctionnalités tactiles sont déjà testées et fonctionnelles, néanmoins, le dialogue par règles et le dialogue libre doivent être testés, ainsi que la synchronisation entre les deux modes d'interaction.

3.5 Conception des tests de validation des fonctions de l'application Amelis

La validation de l'application Amelis se déroule en deux parties. La première partie technique teste le bon déroulement du code et des fonctionnalités créées ; la deuxième partie est expérimentale par rapport aux rencontres avec des utilisateurs

3.5. CONCEPTION DES TESTS DE VALIDATION DES FONCTIONS DE L'APPLICATION AMELIS

pour observer une utilisation réelle du calendrier et le comportement des utilisateurs face au calendrier. Les tests techniques ainsi que les validations auprès des utilisateurs se séparent également en deux parties, l'une sur le dialogue par règles et l'une sur le dialogue libre.

3.5.1 Utilisation de tests end-to-end (E2E) pour le dialogue par règles

Il est expliqué dans le chapitre 2 sur la méthodologie que les tests E2E permettent de simuler des appuis tactiles sur l'écran et de vérifier que les retours visuels sont bien présents. Le dialogue libre ne peut être testé par ce type de tests car on cherche à avoir un événement déjà défini, or la réponse du dialogue libre est aléatoire.

3.5.1.1 Tests sur les réponses de l'agent virtuel

Dans le dialogue par règles, des structures sont conçues pour répondre à certaines demandes de l'utilisateur. Lorsqu'une règle s'active, une réponse prédéfinie doit être donnée par l'agent virtuel. Les tests E2E permettent premièrement de savoir si une règle s'active et deuxièmement si c'est celle attendue. Les règles réagissent à certains mots, à des synonymes, à des dates ou à des heures. Plusieurs tests sont spécifiques à chaque règle. Les tests comprennent donc une formulation différente pour une même demande pour savoir si une règle s'active bien dans les cas souhaités. Il est ainsi testé que les règles s'activent correctement pour une demande qui y est associée, et que les règles ne s'activent pas si la demande n'y correspond pas. Le test est validé si la réponse reçue correspond à la réponse définie dans la règle à vérifier.

3.5.1.2 Tests sur la synchronisation du vocal vers le tactile

Il est évoqué précédemment que chaque action vocale est associée à une fonctionnalité tactile. Cette association implique qu'il y a un retour visuel sur l'interface pour une commande vocale. Les retours visuels sont déjà connus. Lorsqu'un test est lancé, une commande vocale est envoyée au dialogue par règles et le test est validé si le bon composant apparaît ou est modifié à l'écran. Il est également vérifié qu'il n'y a pas d'autres composants qui s'activent ou se modifient.

3.5.1.3 Tests sur la synchronisation du tactile vers le vocal

Comme pour la synchronisation du vocal vers le tactile, à une action tactile est associé un retour vocal. Cependant, plusieurs tests sont effectués :

- Un premier test consiste à activer l’agent virtuel, puis à réaliser une action tactile à laquelle est normalement associée une réaction vocale. Pour ce test, il est vérifié que la réponse de l’agent virtuel correspond à la règle qui s’active par rapport à l’activation de la fonctionnalité tactile à laquelle elle est liée ;
- Un second test consiste à laisser l’agent virtuel inactif lors d’une action tactile. Il est vérifié qu’aucun retour vocal ne s’active. Si cela est validé, l’agent virtuel est activé et il doit donner la réponse associée à l’action tactile réalisée puisque le tactile et le vocal s’échangent sans cesse des messages pour être rendus au même niveau de traitement des demandes ;

La figure 3.14 donne un exemple de test réalisé. Dans cet exemple, le début d’un ajout de rendez-vous est testé.

Lorsque la demande d’ajout est envoyée, la règle devant réagir à l’ajout doit s’activer. La synchronisation du tactile vers le vocal prend place ici lors de l’activation de l’agent virtuel. En activant de façon tactile l’agent virtuel, il doit saluer l’utilisateur. Enfin, la synchronisation du vocal vers le tactile est testée lorsque la demande vocale d’ajout est réalisée. Le panneau d’ajout de rendez-vous doit s’afficher à l’écran.

Les tests E2E permettent de valider le bon fonctionnement technique du dialogue par règles et de la synchronisation entre les deux modes. Le dialogue libre est testé autrement puisqu’il faut vérifier la cohérence des réponses qui sont données suite à un apprentissage et que l’on ne peut pas prédire à l’avance.

3.5.2 Évaluation des apprentissages des réseaux de neurones

Plusieurs réseaux de neurones sont entraînés avec différents paramètres pour trouver l’apprentissage le plus adapté. Un premier test est effectué pour trier ces différents apprentissages.

3.5. CONCEPTION DES TESTS DE VALIDATION DES FONCTIONS DE L'APPLICATION AMELIS

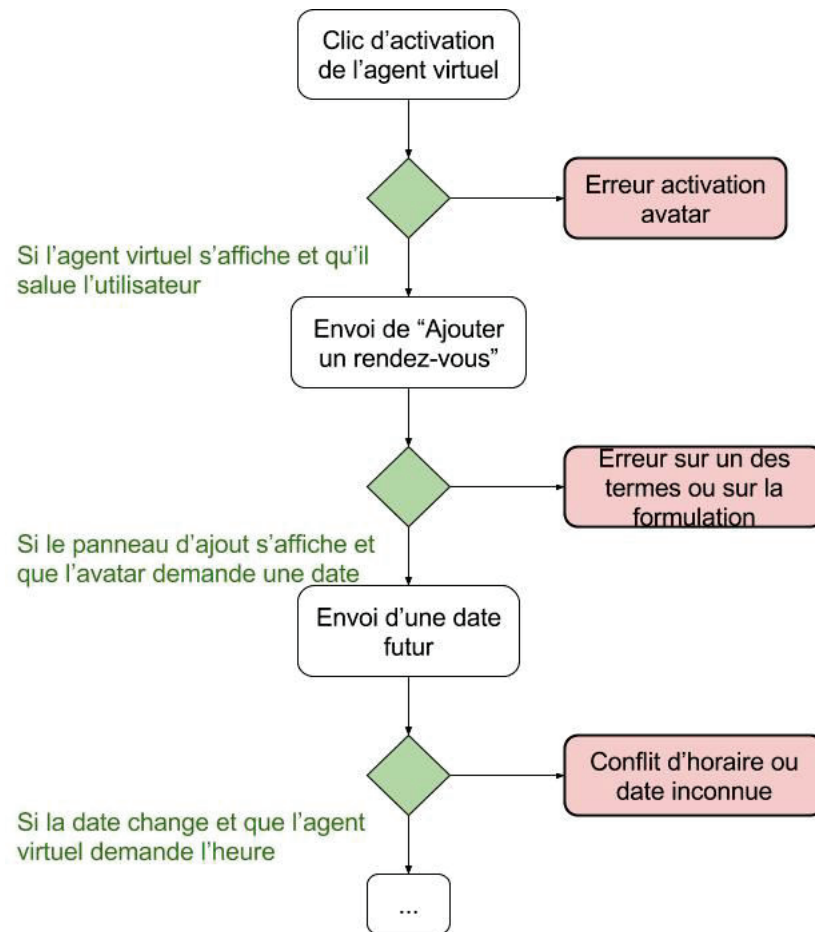


figure 3.14 – Exemple de vérification du dialogue par règles et de la synchronisation

3.5.2.1 Présélection d'apprentissages

Lors d'un apprentissage, deux points permettent de comparer des apprentissages à chaque étape n définie :

- Le réseau de neurones utilise un petit corpus qui ne sert pas pour l'apprentissage pour vérifier l'évolution des réponses de l'apprentissage envers de nouvelles phrases ;
- Le calcul du taux de perte, représentant un calcul d'écart type entre les réponses fournis par l'apprentissage et les réponses du corpus. Par exemple, le corpus comprend une question X et une réponse Y , le réseau donne une réponse Y' à la question X . Il y a ensuite un calcul d'écart entre la réponse attendue Y et la réponse Y' .

A l'aide de ces deux points, il est possible d'effectuer une présélection sur les apprentissages à conserver. Il est donc testé la différence de qualité entre un apprentissage et un autre en se basant sur ces critères. Si un apprentissage a un taux de perte plus bas qu'un autre à une même étape n , il est supposé meilleur, mais cela est confirmé avec la pertinence des réponses face à de nouvelles questions si les réponses sont plus structurées d'un point de vue grammatical et si elles sont plus proches des réponses attendues aux questions.

3.5.2.2 Test BLEU et test sémantique

Il est expliqué dans le chapitre 2 sur la méthodologie l'apport d'un test BLEU dans l'évaluation d'un dialogue. Il est possible de trouver des modèles libres d'utilisation déjà existants permettant d'effectuer ce test [98]. Comme pour le test BLEU, il est possible de trouver des modèles libres d'accès pour effectuer un test sémantique [99]. Les deux tests sont effectués sur les apprentissages retenus lors de la présélection, permettant ainsi de choisir le meilleur apprentissage.

Suite à ces tests techniques, un apprentissage est choisi permettant d'avoir un dialogue libre. L'application Amelis est donc testée. Pour valider l'application Amelis, deux types de résultats vont être étudiés. Les résultats sur les tests techniques et les résultats avec les utilisateurs. Cela permet d'avoir une validation technique de l'application mais aussi une validation en utilisation réelle face à de réelles attentes.

Chapitre 4

Résultats

4.1 Résultats aux tests techniques

Pour chaque objectif, plusieurs hypothèses sont émises. Certaines sont validées ou invalidées par les tests techniques et d'autres par la validation avec les utilisateurs. Cette partie présente les résultats obtenus avec les tests techniques. Elle se décompose en deux parties, la première réalisée avec les tests E2E, qui sont des tests consacrés aux tests pour le dialogue par règles, la deuxième avec les tests BLEU et sémantiques qui sont des tests liés à l'évaluation du dialogue libre et plus précisément de l'apprentissage lié au réseau de neurones.

4.1.1 Les tests E2E

Les tests E2E ont évolué et ont augmenté à chaque nouvelle fonctionnalité ajoutée pour le vocal ou pour la synchronisation entre le vocal et le tactile. Le module de test comprend 638 tests. Trois quart de ces tests sont consacrés aux commandes liées aux rendez-vous. Le quart restant est quant à lui consacré à la navigation entre les pages en vocal pour vérifier que suite à diverses commandes le calendrier affiche la bonne information. Les commandes peuvent être : "Affiche moi les photos" ou des demandes liées à la météo et à la date du jour, telles que "Quel jour est-on?" ou "Quelle heure est-il?". Durant la conception des tests, lorsque des erreurs sont

identifiées, un message explique comment le test n'a pas fonctionné et quelle partie n'a pas fonctionné. Comme illustré dans la figure 4.1, un message d'erreur est présent. Il est marqué que c'est un test lié à la consultation de rendez-vous et que le test a pour but d'afficher un rendez-vous parmi plusieurs rendez-vous présents dans la même journée mais à des heures différentes. L'erreur explique que l'agent virtuel aurait dû trouver le rendez-vous et répondre "Voici le rendez-vous". Cette erreur, il est possible de remonter au test associé. Étant donné que la demande envoyée à l'agent virtuel est "Je veux consulter mon rendez-vous aujourd'hui à 15h" et qu'il y a un rendez-vous à 15h et un à 17h, le fait que l'agent virtuel ait répondu "Je ne trouve pas de rendez-vous à cette date. Pouvez-vous me donner une autre date?" implique que la règle pour la consultation s'est bien activée mais que la requête pour récupérer le rendez-vous n'a pas fonctionné.

```

Failures:
1) test activation consulter d'un rendez-vous - plusieurs rdv existant même jour
   mais heures dif et bonnes #1
   Message:
     Expected 'Je ne trouve pas de rendez-vous à cette date. Pouvez-vous me donne
r une autre date?' to contain 'Voici le rendez-vous'.

```

figure 4.1 – Exemple d'erreur suite à un test

Les erreurs sont résolues au fur et à mesure avant de lancer de nouveaux tests. La figure 4.2 donne un exemple d'un test sur un ajout et des sous-tests mis en place sous la forme d'un diagramme de langage de modélisation unifié (UML) [100].

1. Le module de test envoie la phrase « Je veux ajouter un rendez-vous », un sous-test vérifie que la phrase apparaît dans la boîte de dialogue avec l'agent virtuel ;
2. Un sous-test vérifie que le panneau d'ajout de rendez-vous s'affiche à l'écran ;
3. Un sous-test vérifie que l'agent virtuel donne une réponse dans la boîte de dialogue et que c'est une demande de date ;
4. Un sous-test vérifie que la nouvelle phrase envoyée apparaît dans la boîte de dialogue avec l'agent virtuel ;
5. Un sous-test vérifie que la date dans le panneau d'ajout de rendez-vous change

4.1. RÉSULTATS AUX TESTS TECHNIQUES

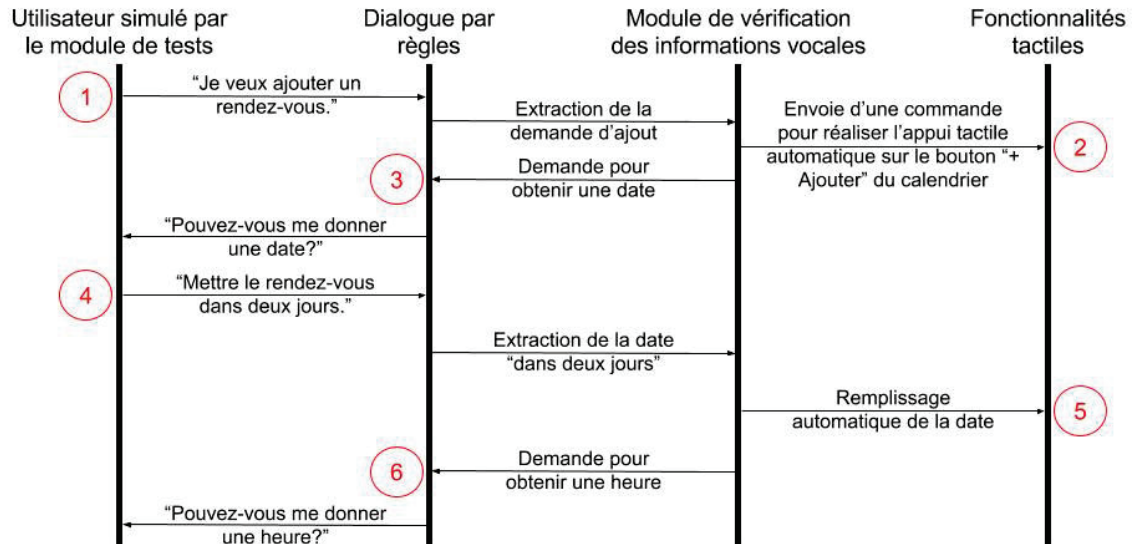


figure 4.2 – Exemple du déroulement d'un test E2E d'ajout sous la forme d'un diagramme UML

et qu'elle correspond à la date souhaitée (ici, il est vérifié que la date est incrémentée de 2 jours) ;

6. Un sous-test vérifie que l'agent virtuel donne une réponse dans la boîte de dialogue et que c'est une demande d'heure.

Tous les tests ont un déroulement similaire, bien que certains soient plus complexes que d'autres. Pour valider d'un point de vue technique l'application Amelis, il a donc fallu refactoriser le code pour que les 638 tests soient validés. Il faut savoir que beaucoup de tests se ressemblent pour une même tâche, comme le montre la figure 4.3. Cela permet de vérifier que plusieurs formulations obtiennent un même résultat. Dans ces tests se trouvent à la fois différentes formulations pour le terme "rendez-vous" et pour le verbe "ajouter", pour valider qu'une règle réagit à divers mots et à divers ordres de ces mots. Par exemple, "Je veux ajouter un rendez-vous demain" et "Pour demain, j'aimerais ajouter un rendez-vous".

Dans la figure 4.3, un ajout simple correspond à un ajout sans autre information. Chaque test est séparé selon les informations à envoyer. Par exemple, il y a des tests pour l'ajout sans information et il y a des tests pour l'ajout avec une date fournie dès la première demande de l'utilisateur.


```
it("#4", function() {
  avatar.ask("je veux mettre un rendez vous");
  expect(avatar.answers.last().getText()).toContain(reponse);
});

it("#7", function() {
  avatar.ask("j'ai un appointment à ajouter");
  expect(avatar.answers.last().getText()).toContain(reponse);
});

it("#8", function() {
  avatar.ask("peux tu ajouter un rendez vous ?");
  expect(avatar.answers.last().getText()).toContain(reponse);
});

it("#9", function() {
  avatar.ask("je voudrais rajouter un rendez vous");
  expect(avatar.answers.last().getText()).toContain(reponse);
});

it("#11", function() {
  avatar.ask("ajouter activité");
  expect(avatar.answers.last().getText()).toContain(reponse);
});

it("#12", function() {
  avatar.ask("je veux créer un rendez vous");
  expect(avatar.answers.last().getText()).toContain(reponse);
});
```

figure 4.3 – Exemple de tests sur un ajout simple de rendez-vous

4.2. RÉSULTATS DE LA PREMIÈRE EXPÉRIMENTATION

4.1.2 Le test BLEU et le test sémantique

Deux modèles sont retenus suite à différents entraînements. Cependant, un seul est testé car sans effectuer de test technique. Il faut noter que l'apprentissage d'un réseau de neurones prend du temps. Par exemple, pour le modèle MNIST [101], qui est un modèle apprenant à reconnaître un chiffre à partir d'un corpus de plusieurs images représentant toutes un chiffre, l'entraînement demande plusieurs heures, or ce modèle est considéré comme un modèle simple à entraîner. Pour un modèle "sequence to sequence" (seq2seq), la durée d'apprentissage pour un large modèle, c'est à dire sur un corpus d'environ 200000-300000 phrases, il est estimé qu'en 2-3 jours et avec 8 GPUs TitanX (processeur graphique) l'entraînement serait terminé. Pour nos apprentissage, nous ne disposons que d'un GPU de base et notre corpus comprend 250000 phrases. Les deux apprentissages ne sont donc pas des apprentissages finis par manque de temps. Le test BLEU et le test sémantique ne sont réalisés que sur un apprentissage car l'autre n'a pas encore une syntaxe correcte. L'apprentissage retenu est celui de Google. La syntaxe de cet apprentissage est correcte, comme le montreront les extraits de conversations présents dans les résultats de la deuxième expérimentation. Le score BLEU obtenu est de 2%, ce qui se situe dans la zone où la qualité de réponse n'est pas considérée comme compréhensible [91]. Pour la similarité des réponses supposées attendues et celles obtenues d'un point de vue sémantique, le score obtenu est de 9%.

4.2 Résultats de la première expérimentation

La première expérimentation porte sur l'utilisation d'interactions tactiles et vocales. L'utilisation du vocal se concentre ici sur le dialogue par règles. Le premier but est d'observer l'utilisation faite uniquement par le dialogue par règles, de comprendre les erreurs liées au vocal et d'identifier les points manquants ou à améliorer du dialogue par règles. Le second but est d'observer si les utilisateurs privilégient un mode d'interaction, soit le tactile, soit le vocal et si oui d'identifier pourquoi un mode est plus utilisé/apprécié qu'un autre.

Les résultats présentés ici se rapportent principalement à l'utilisation du dialogue par règles. Il est également comparé l'utilisation du tactile et du vocal faites tout au

long de l'expérimentation, qui dure 3 mois. Les résultats sont développés dans l'ordre suivant :

- L'évaluation cognitive des participants d'après un test appelé "Montreal Cognitive Assessment" (MoCA) ;
- L'utilisation du dialogue par règles ;
- Les causes des erreurs vocales ;
- L'utilisation du tactile en comparaison de l'utilisation vocale au cours de la semaine d'expérimentation. Chaque participant est identifié par un numéro Pi-Ej où i correspond au numéro du participant, entre 1 et 4, et j correspond au numéro de l'expérimentation entre 1 et 2. ;
- L'utilisation du dialogue libre.

4.2.1 Présentation des participants

Les participants sont âgées entre 70 et 90 ans. Les figures 4.4, 4.5, 4.6, 4.7 reprennent les informations essentielles de chaque participant, ainsi que le résultat de leur test cognitif. Il est rappelé qu'un score cognitif égal ou supérieur à 26 sur 30 indique que la personne ne présente pas de troubles cognitifs.

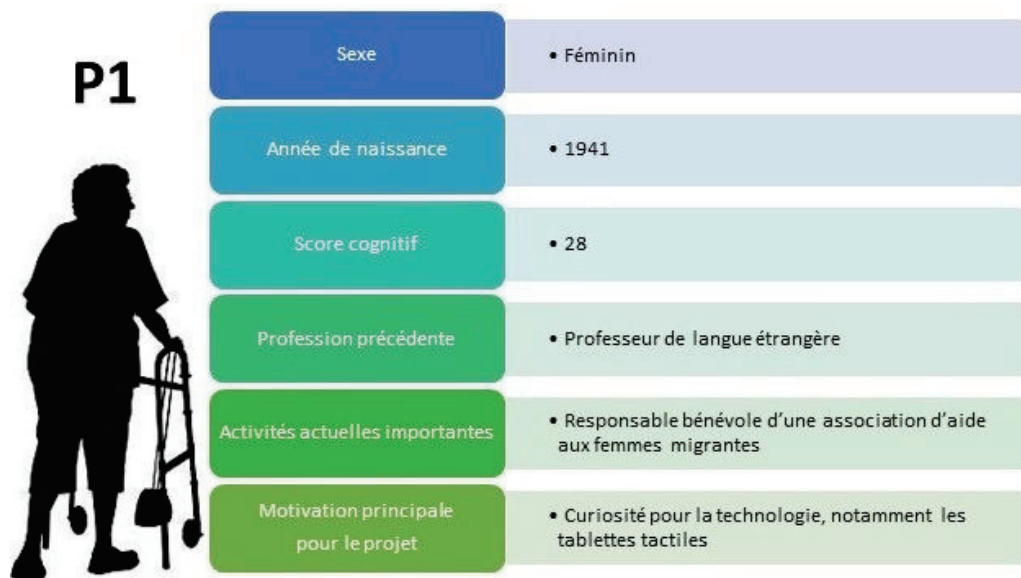


figure 4.4 – Présentation de la participante P1_E1

4.2. RÉSULTATS DE LA PREMIÈRE EXPÉRIMENTATION



figure 4.5 – Présentation de la participante P2_E1

Chaque participant a également passé un test de reconnaissance vocale, comme expliqué dans la méthodologie 2.4 du chapitre 2. Les scores sont résumés ci-dessous :

- La participante P1_E1 a obtenu un taux de reconnaissance vocale de 31% de réussite pour les textes sans pause et 73% de réussite pour les textes avec des pauses ;
- La participante P2_E1 a obtenu un taux de reconnaissance vocale de 75% de réussite pour les textes sans pause et 76% de réussite pour les textes avec des pauses ;
- La participante P3_E1 a obtenu un taux de reconnaissance vocale de 59% de réussite pour les textes sans pause et 76% de réussite pour les textes avec des pauses ;
- Le participant P4_E1 a obtenu un taux de reconnaissance vocale de 15% de réussite pour les textes sans pause et 71% de réussite pour les textes avec des pauses ;

Les résultats de chaque participant sont présentés pour chacun des résultats. Cela permet d'analyser le comportement de chaque utilisateur face à l'application Amelis.



figure 4.6 – Présentation de la participante P3_E1



figure 4.7 – Présentation du participant P4_E1

4.2. RÉSULTATS DE LA PREMIÈRE EXPÉRIMENTATION

4.2.2 Taux de réussite des actions vocales des participants

Pour une meilleure lisibilité, les résultats sont présentés sous forme graphique. Les quatre figures suivantes comparent le nombre d'actions vocales totales effectuées au nombre d'actions réussies lors d'une utilisation d'Amelis pendant 3 mois.

La figure 4.8 permet de voir que la participante P1_E1 a beaucoup utilisé le vocal pour ajouter des rendez-vous. Elle a également communiqué de manière libre avec l'agent virtuel, mais sans le dialogue libre par réseaux de neurones qui n'était pas compatible lors de la première expérimentation.

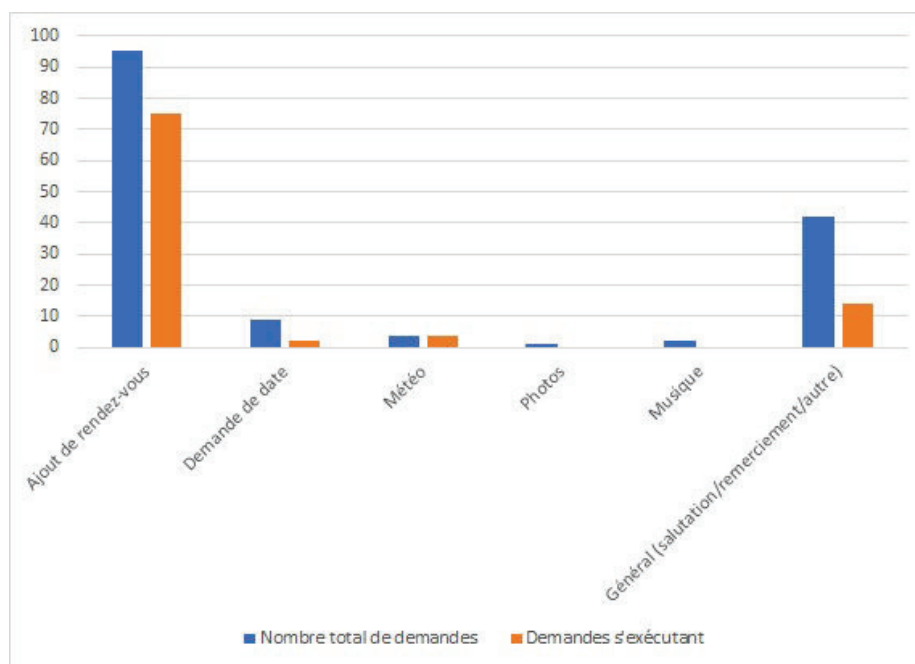


figure 4.8 – Nombre d'actions vocales réussies de la participante P1_E1

La figure 4.9 montre que la participante P2_E1 a très peu utilisé le vocal.

La figure 4.10 montre que la participante P3_E1 a réalisé une utilisation variée des fonctionnalités vocales. De plus, l'utilisation du vocal est plus centrée que l'ajout de rendez-vous et la discussion libre.

Dans la figure 4.11 est visible une utilisation du vocal grandement centrée sur les rendez-vous et en particulier l'ajout.

Cela n'est pas représenté dans les figures précédentes, mais les utilisateurs ont

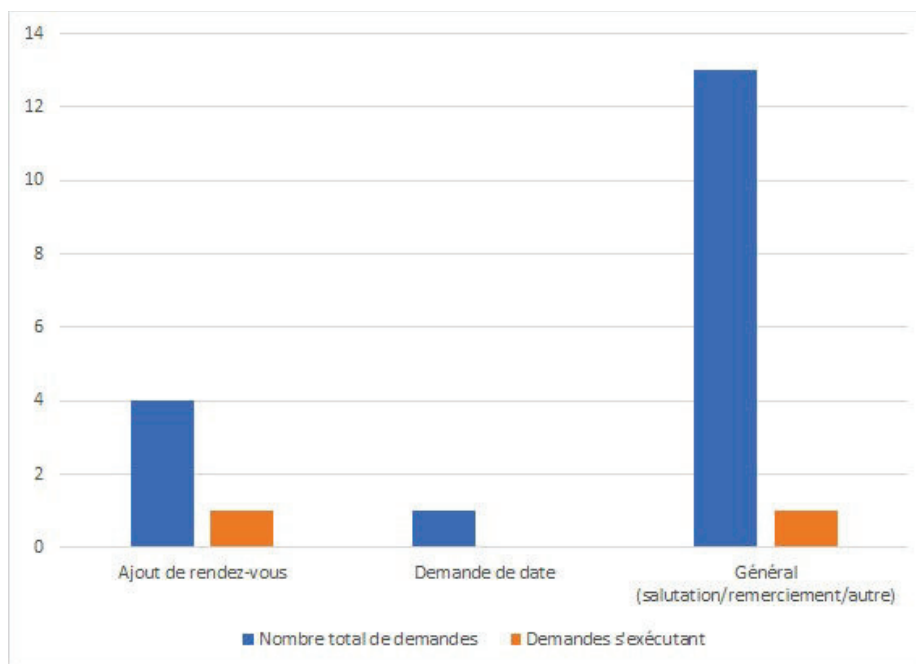


figure 4.9 – Nombre d'actions vocales réussies de la participante P2_E1

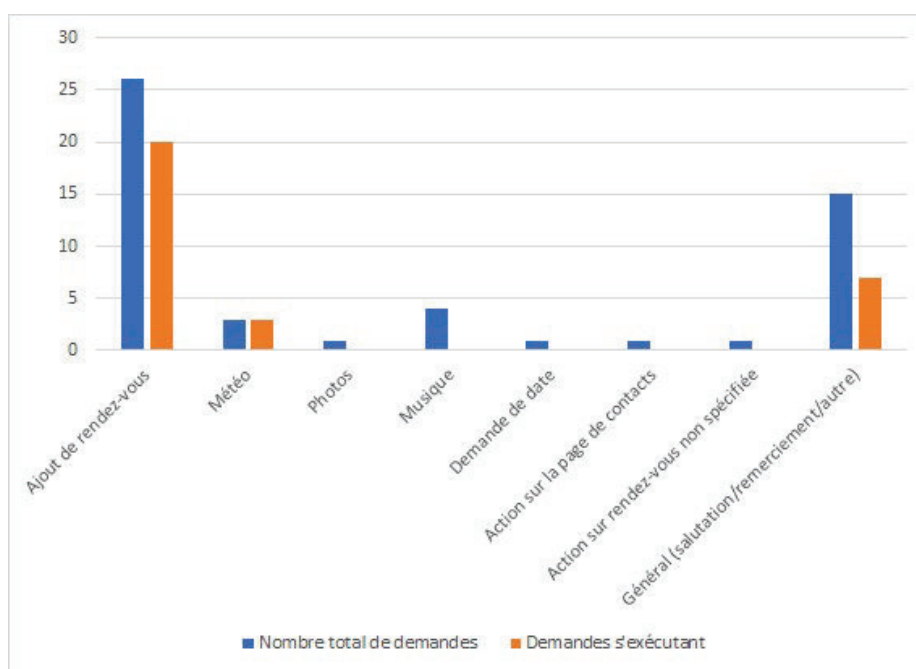


figure 4.10 – Nombre d'actions vocales réussies de la participante P3_E1

4.2. RÉSULTATS DE LA PREMIÈRE EXPÉRIMENTATION

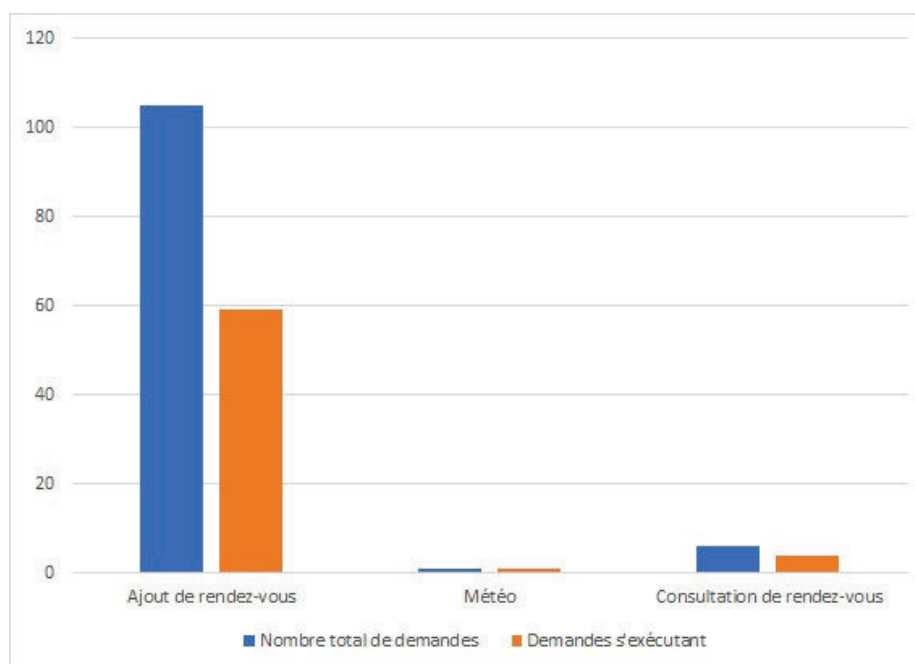


figure 4.11 – Nombre d’actions vocales réussies du participant P4_E1

enregistré des rendez-vous en plein milieu de la nuit (entre 23h et 4h du matin) :

- P1_E1 en a enregistré 10 ;
- P2_E1 en a enregistré 1 ;
- P3_E1 en a enregistré 1 ;
- P4_E1 en a enregistré 7.

Pour chaque participant, à part la deuxième qui a très peu utilisé le vocal, il est constaté que le taux d’erreurs vocales varie entre 38 et 45%. La participante P2_E1 est quant à elle à 88% de taux d’erreurs vocales, mais elle s’est peu servie du vocal et elle l’a principalement utilisé pour le dialogue libre qui est, lors de cette expérimentation, traité par le dialogue par règles. Il n’est donc pas approprié pour traiter un dialogue libre. Les erreurs vocales se décomposent en plusieurs causes.

4.2.3 Causes des erreurs vocales

La figure 4.12 présente les causes des erreurs vocales de la participante P1_E1. Environ la moitié des erreurs vocales sont dues à une formulation incomplète de la part

de cette utilisatrice. Les deux autres causes d'erreurs vocales les plus importantes sont la reconnaissance vocale et le manque de règles traitant les fonctionnalités. Il arrive aussi que la participante ne réponde pas au sujet demandé par l'agent virtuel ce qui mène à une erreur vocale. Pour cette participante :

- Un manque au niveau des règles représente 27% des erreurs vocales ;
- Les formulations incomplètes représentent 40% des erreurs vocales ;
- La reconnaissance vocale représente 30% des erreurs vocales ;
- Les réponses hors sujet représentent 3% des erreurs vocales.

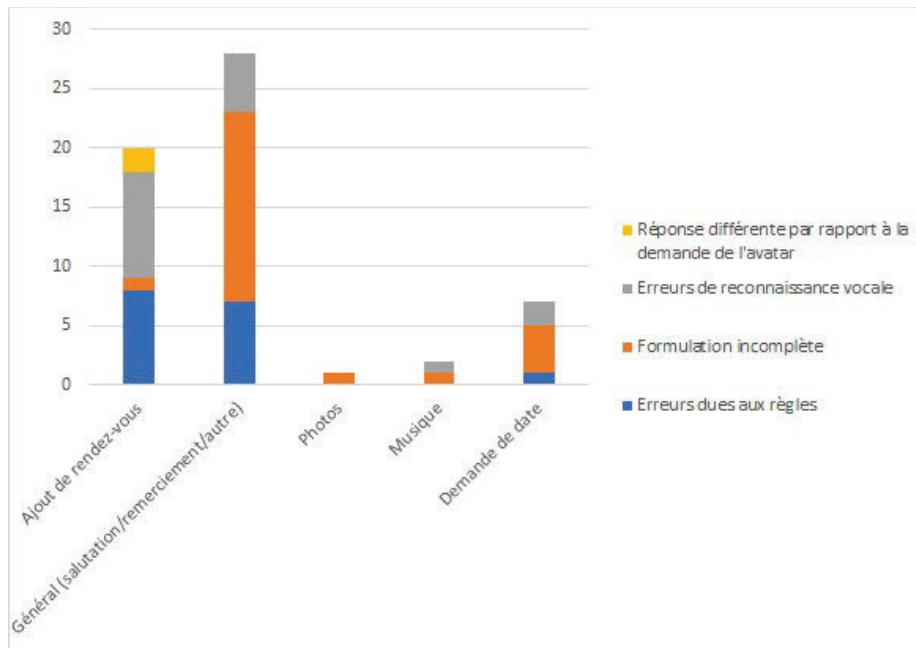


figure 4.12 – Causes des erreurs vocales de la participante P1_E1

La figure 4.13 montre les erreurs vocales de la participante P1_E2. Elles sont majoritairement dues à un manque au niveau des règles. Il est toutefois nécessaire de rappeler que la participante P2_E1 a peu utilisé le vocal. Pour cette participante :

- Un manque au niveau des règles représente 87% des erreurs vocales ;
- Les formulations incomplètes représentent 19% des erreurs vocales ;
- La reconnaissance vocale représente 4% des erreurs vocales.

La figure 4.14 montre également que les erreurs vocales pour la participante P3_E1 sont principalement dues à un manque de règles. Des erreurs sont également dues

4.2. RÉSULTATS DE LA PREMIÈRE EXPÉRIMENTATION

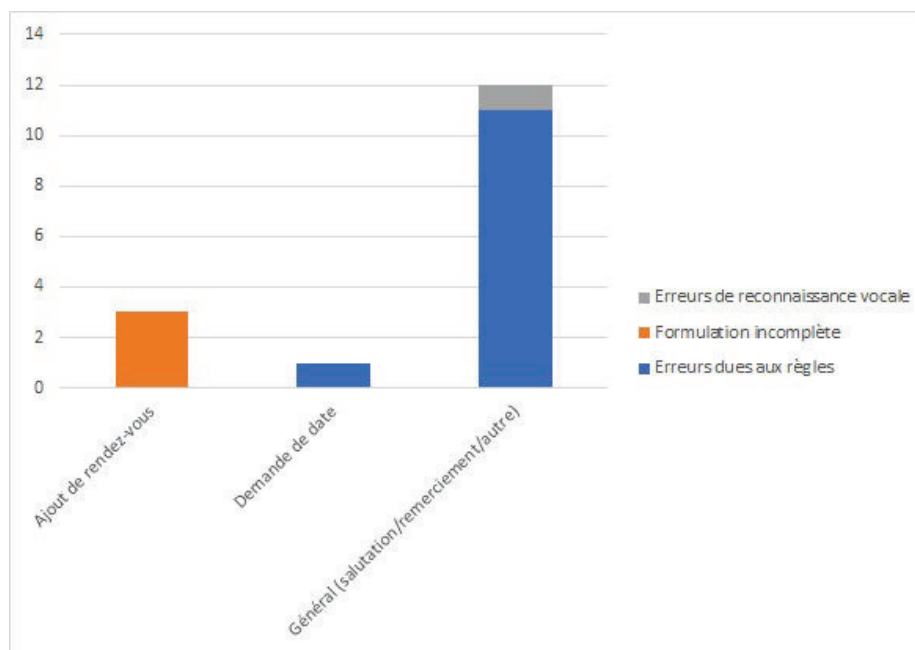


figure 4.13 – Causes des erreurs vocales de la participante P2_E1

à un problème de reconnaissance vocale et de formulation incomplètes. Pour cette participante :

- Un manque au niveau des règles représente 72% des erreurs vocales ;
- Les formulations incomplètes représentent 14% des erreurs vocales ;
- La reconnaissance vocale représente 14% des erreurs vocales.

Dans la figure 4.15 des causes des erreurs vocales du participant P1_E4, un fort taux d'erreurs vocales proviennent d'un manque de règles, mais aussi de la reconnaissance vocale. D'autres erreurs sont dues à une formulation incomplète et à des réponses hors sujet de l'utilisateur. Pour ce participant :

- Un manque au niveau des règles représente 49% des erreurs vocales ;
- Les formulations incomplètes représentent 9% des erreurs vocales ;
- La reconnaissance vocale représente 29% des erreurs vocales ;
- Les réponses hors sujet représentent 13% des erreurs vocales ;

Les figures 4.8, 4.9, 4.10, 4.11 montre une utilisation globale du vocal. Chaque participant a également utilisé le tactile.

CHAPITRE 4. RÉSULTATS

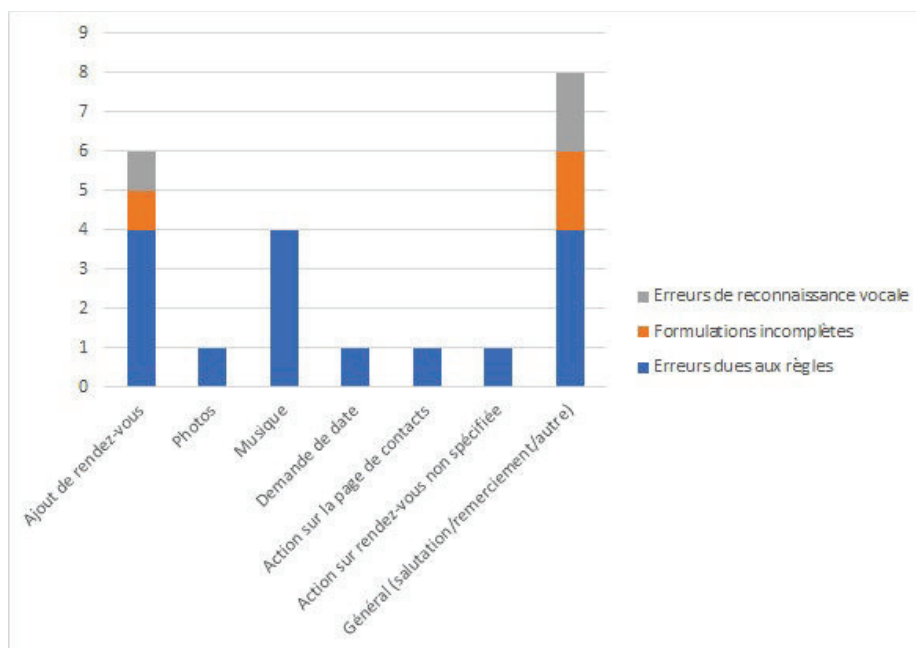


figure 4.14 – Causes des erreurs vocales de la participante P3_E1

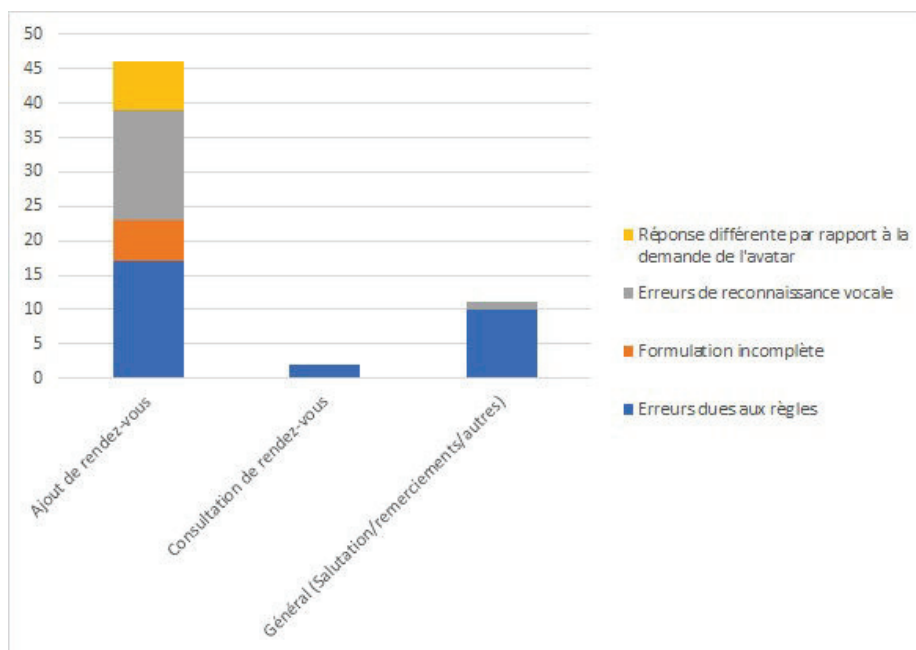


figure 4.15 – Causes des erreurs vocales du participant P4_E1

4.2. RÉSULTATS DE LA PREMIÈRE EXPÉRIMENTATION

4.2.4 Utilisation tactile et vocale durant les 3 mois d'expérimentation

L'utilisation de l'application Amelis est décomposée en deux parties, une pour résumer la fréquence d'utilisation du tactile et du vocal, et une autre sur l'utilisation de chaque fonctionnalité au cours des semaines.

4.2.4.1 Fréquence d'utilisation du tactile et du vocal

Les quatre tableaux suivants résument la fréquence d'utilisation globale du tactile et du vocal. Les tableaux reprennent le nombre de jours où le tactile ou le vocal sont utilisés et non les nombres d'actions tactiles ou vocales. Dans le tableau 4.1, il est visible que la participante P1_E1 a une utilisation plus fréquente du tactile. De plus, l'utilisation du vocal a presque disparu au cours du dernier mois.

tableau 4.1 – Fréquence d'utilisation tactile et vocale au cours des 3 mois d'expérimentation de la participante P1_E1

	Premier mois	Deuxième mois	Troisième mois
Nombre de jours avec utilisation du vocal	6	7	2
Nombre de jours avec utilisation du tactile	15	30	19

Le tableau 4.2 rappelle que la participante P2_E1 a peu utilisé le vocal. Néanmoins, la participante s'est beaucoup plus servie du tactile pour interagir avec l'application Amelis.

tableau 4.2 – Fréquence d'utilisation tactile et vocale au cours des 3 mois d'expérimentation de la participante P2_E1

	Premier mois	Deuxième mois	Troisième mois
Nombre de jours avec utilisation du vocal	3	2	0
Nombre de jours avec utilisation du tactile	9	15	1

Pour la participante P3_E1, la fréquence d'utilisation du tactile et du vocal sont

à peu près similaires à la participante P2_E1 bien qu'elles soient légèrement supérieures.

tableau 4.3 – Fréquence d'utilisation tactile et vocale au cours des 3 mois d'expérimentation de la participante P3_E1

	Premier mois	Deuxième mois	Troisième mois
Nombre de jours avec utilisation du vocal	4	5	0
Nombre de jours avec utilisation du tactile	12	16	6

Pour le participant P4_E1, il est intéressant de voir qu'il y a une forte fréquence d'utilisation du tactile tout au long des 3 mois. À l'inverse, l'utilisation a diminué pour quasiment disparaître le dernier mois.

tableau 4.4 – Fréquence d'utilisation tactile et vocale au cours des 3 mois d'expérimentation du participant P4_E1

	Premier mois	Deuxième mois	Troisième mois
Nombre de jours avec utilisation du vocal	11	8	1
Nombre de jours avec utilisation du tactile	19	31	26

Les tableaux précédents résument la fréquence d'utilisation tactile et du vocal. Le prochain paragraphe montre une utilisation plus détaillée selon chacune des fonctionnalités de l'application Amelis.

4.2.4.2 Utilisation du tactile et du vocal selon les fonctionnalités de l'application Amelis

Les figures qui suivent reprennent l'utilisation de toutes les fonctionnalités qui sont partagées par le tactile et le vocal. Par exemple, il n'y a pas la météo, car aucune fonctionnalité tactile n'est liée à la météo, tandis qu'il est possible de la demander à l'aide du vocal. Pour une meilleure visualisation des utilisations effectuées, il est décidé de présenter les résultats au fil des semaines. De plus, il y a un fort contraste dans les échelles entre l'utilisation tactile et vocale car pour faire une tâche, l'interaction

4.3. RÉSULTATS DE LA DEUXIÈME EXPÉRIMENTATION

vocale demande beaucoup moins d'actions. Par exemple, pour remplir le titre d'un rendez-vous, il suffit de donner le titre en vocal comme "Tournoi de bingo" ce qui prend une action, tandis qu'en tactile, dans le titre "Tournoi de bingo", il y a 17 caractères à rentrer et chaque appui représente une action.

La figure 4.17 montre une utilisation plus importante dans les premières semaines de la part de la participante P1_E1 qui est une utilisation principalement tactile. Il n'y a que pour l'ajout de rendez-vous où le vocal est utilisé de manière constante. Le reste des fonctionnalités est utilisé majoritairement en tactile.

Pour la participante P2_E1, l'utilisation de l'application Amelis qui est principalement tactile se décompose par pics d'utilisation. Par exemple, pour la participante P1_E1, l'utilisation est plus continue. Les actions tactiles sont principalement centrées autour des rendez-vous ou des contacts.

Dans la figure 4.18, la participante P3_E1 n'a utilisé l'application Amelis que dans la première moitié des 3 mois d'expérimentation. L'utilisation est partagée entre chaque fonctionnalité bien qu'elle soit plus importante pour la musique, les contacts, les paramètres et l'ajout de rendez-vous. L'utilisation vocale est à peine visible car elle est faible en la répartissant sur les semaines et la différence d'échelle entre le tactile et le vocal est grande. Par exemple, pour l'ajout de rendez-vous, la courbe du vocal reste le long de l'axe des abscisses alors l'utilisation vocale est la plus importante.

Pour le participant P4_E1, il y a une très forte utilisation du tactile au cours du temps pour les actions sur les rendez-vous. Pour le vocal, il est visible que l'utilisation est centrée autour des rendez-vous mais qu'elle cesse au bout de la 7ème semaine d'expérimentation.

4.3 Résultats de la deuxième expérimentation

Les résultats présentés ici se rapportent principalement à l'utilisation de la synchronisation entre les interactions tactiles et vocales. Il est également montré l'utilisation du dialogue libre. Pour rappel, la deuxième expérimentation a duré une semaine. Elle se décompose entre des rencontres à domicile avec les participants pour leur demander d'effectuer certaines tâches et une utilisation libre de l'application Amelis par les participants durant 4 jours à leur domicile. Lors des rencontres, les participants

CHAPITRE 4. RÉSULTATS



Figure 4.16 – Utilisation du tactile et du vocal selon les fonctionnalités de l'application Amelis de la participante P1_E1

4.3. RÉSULTATS DE LA DEUXIÈME EXPÉRIMENTATION

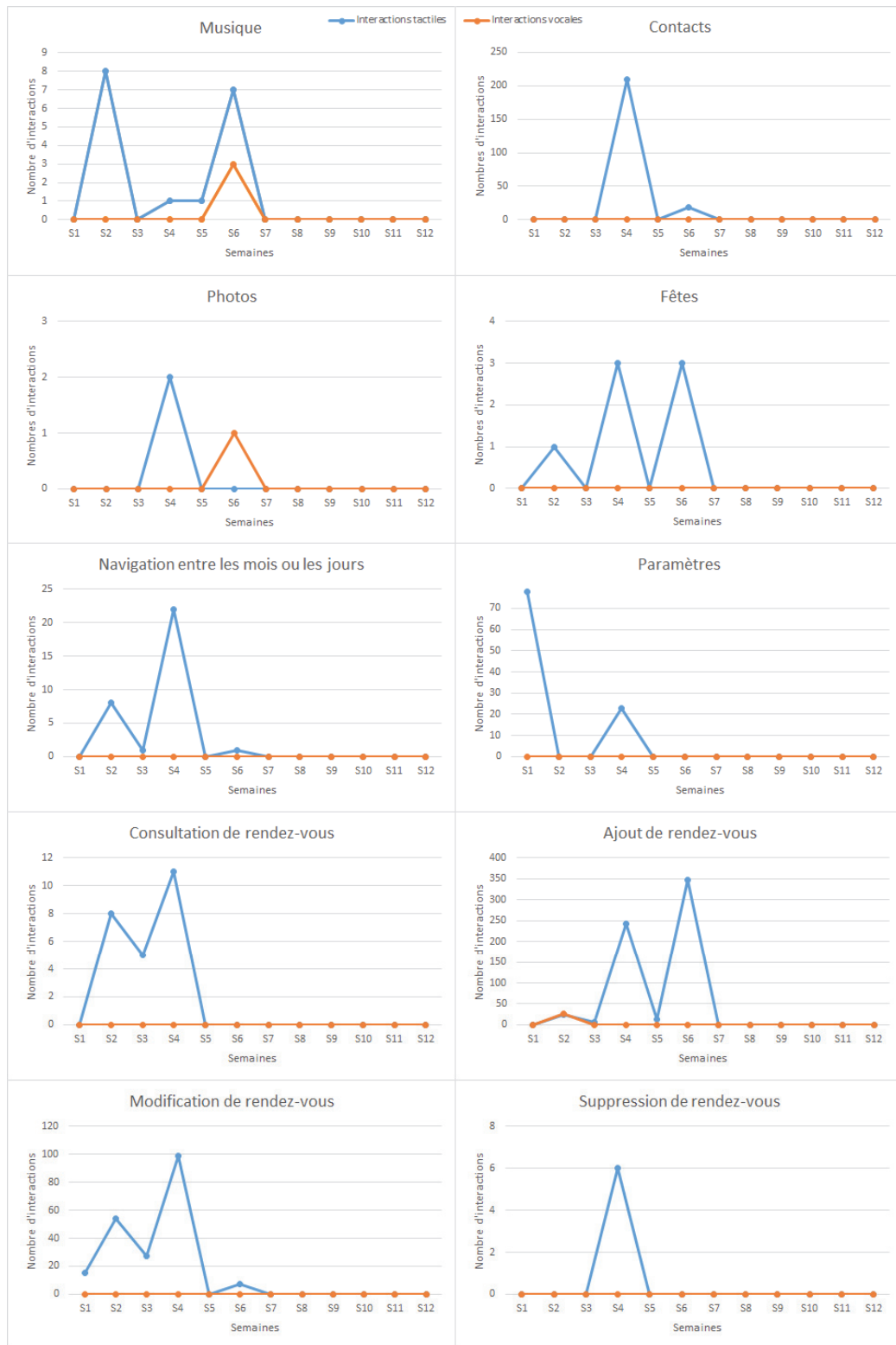


figure 4.17 – Utilisation du tactile et du vocal selon les fonctionnalités de l'application Amelis de la participante P2_E1

CHAPITRE 4. RÉSULTATS



Figure 4.18 – Utilisation du tactile et du vocal selon les fonctionnalités de l'application Amelis de la participante P3_E1

4.3. RÉSULTATS DE LA DEUXIÈME EXPÉRIMENTATION

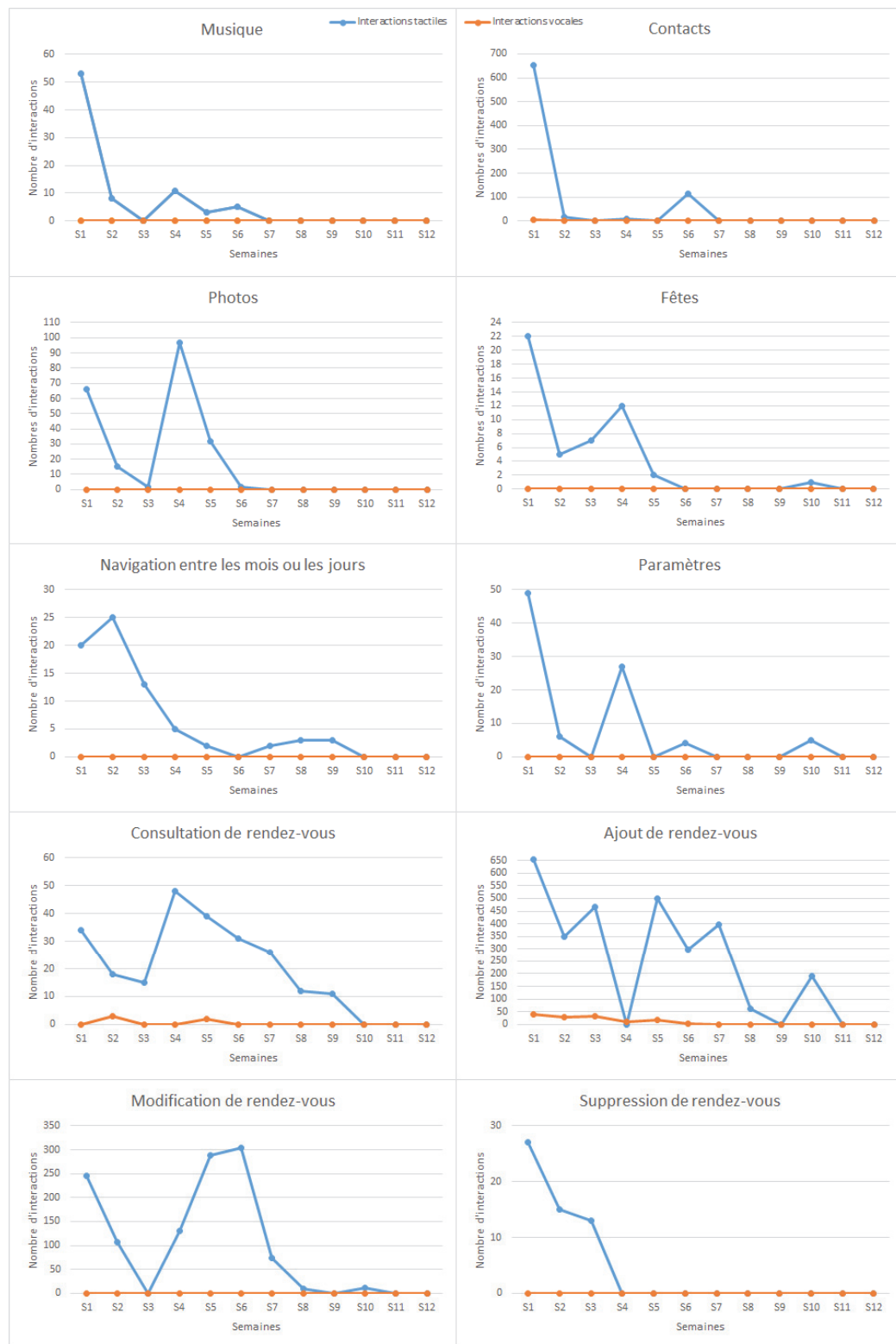


figure 4.19 – Utilisation du tactile et du vocal selon les fonctionnalités de l'application Amelis du participant P4_E1

sont encadrés par un formateur. Ce dernier a pour tâche d'expliquer de manière brève l'application Amelis pour permettre à la fois au participant de savoir ce que comporte l'application, mais aussi de découvrir par lui-même comment utiliser les différentes fonctionnalités. Durant les rencontres, le formateur demande à l'utilisateur de réaliser certaines tâches pour étudier les réactions et la manière dont s'adapte l'utilisateur face à l'application Amelis. Durant les 4 jours où la tablette est laissée au domicile des participants, des rappels sont prévus pour demander à l'utilisateur d'effectuer certaines tâches avec l'application. Les utilisateurs sont libres d'effectuer ces tâches, ainsi que toute autre action. Les résultats sont développés dans l'ordre suivant :

- L'évaluation cognitive des participants d'après un test appelé "Montreal Cognitive Assessment" (MoCA) ;
- La rencontre entre le participant et le formateur ;
- L'utilisation du dialogue par règles et du tactile ;
- Les cas d'utilisation de la synchronisation ;
- L'utilisation du dialogue libre ;

4.3.1 Présentation des participants

Il est noté qu'un participant, le participant P2_E4 a déjà pris part à la première expérimentation. Les autres participants n'ont quant à eux jamais utilisé Amelis.

4.3.2 Résultats des rencontres entre les participants et l'expérimentateur

Durant ces rencontres, il est testé la manière dont l'utilisateur se sert du tactile et du vocal, ainsi que de la synchronisation entre ces deux modes d'interaction. Quatre tâches sont demandées aux participants :

- Un ajout en n'utilisant que le tactile ;
- Un ajout en utilisant le tactile puis le vocal ;
- Un ajout en utilisant le vocal puis le tactile ;
- Un ajout en n'utilisant que le vocal.

Pour chaque tâche, l'utilisateur doit évaluer sa sensation face au(x) mode(s) d'interaction(s) qu'il a utilisé.

4.3. RÉSULTATS DE LA DEUXIÈME EXPÉRIMENTATION

Toutes les évaluations sont reprises dans les figures 4.20, 4.21, 4.22 et 4.23

Le critère de facilité correspond à la facilité de compréhension de la tâche, l'intuitivité correspond au côté intuitif de l'interaction que l'utilisateur réalise, la simplicité correspond à la simplicité d'exécution de la tâche, l'ergonomie correspond à la simplicité de compréhension et à la disposition des composants visuels et des retours à l'écran, enfin la satisfaction correspond au ressenti face à la réalisation de la tâche.

Comme le montre la figure 4.20, la participante P1_E2 n'a pas apprécié l'utilisation de la synchronisation du vocal vers le tactile. Elle a également des difficultés pour la tâche uniquement tactile. A l'inverse, elle est satisfaite de l'interaction simplement vocale et de la synchronisation lorsqu'il s'agit de passer du tactile au vocal.

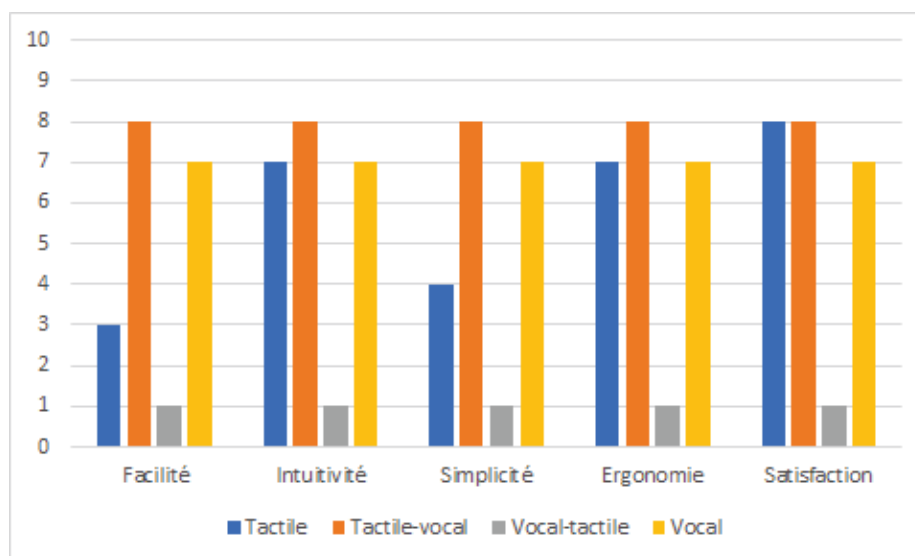


figure 4.20 – Évaluation des quatre tâches d'ajout par l'utilisatrice P1_E2

Dans la figure 4.21, l'utilisatrice P2_E2 est globalement satisfaite et à l'aise avec les différents modes d'interaction et de synchronisation dans les deux sens.

Pour la participante P3_E2, l'interaction tactile n'est pas du tout appréciée. De plus, elle n'est pas à l'aise avec les différents modes de synchronisation. A l'inverse, elle est fortement satisfaite de l'interaction vocale pour l'ajout de rendez-vous.

Pour le participant P4_E2, le ressenti est mitigé par rapport aux différentes interactions. À part pour la synchronisation du tactile vers le vocal qui n'est pas du tout apprécié.

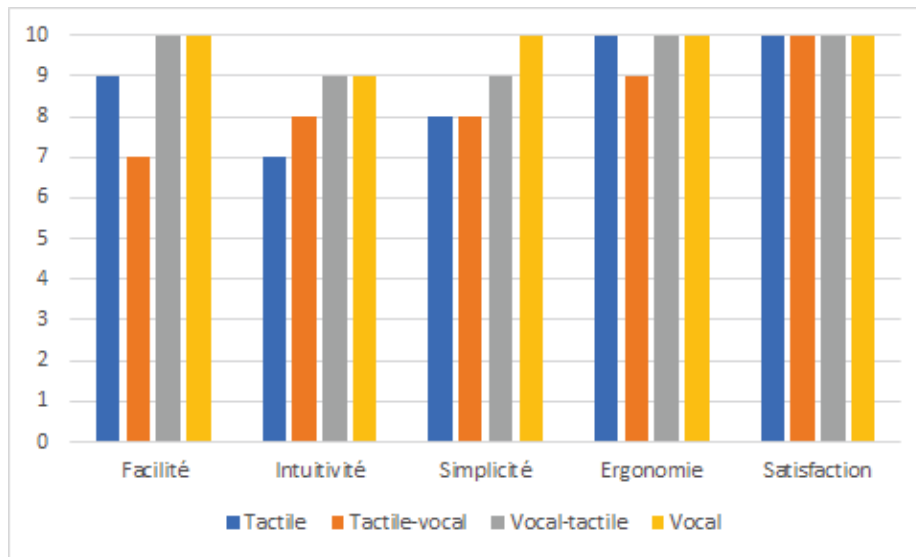


figure 4.21 – Évaluation des quatre tâches d’ajout par l’utilisatrice P2_E2

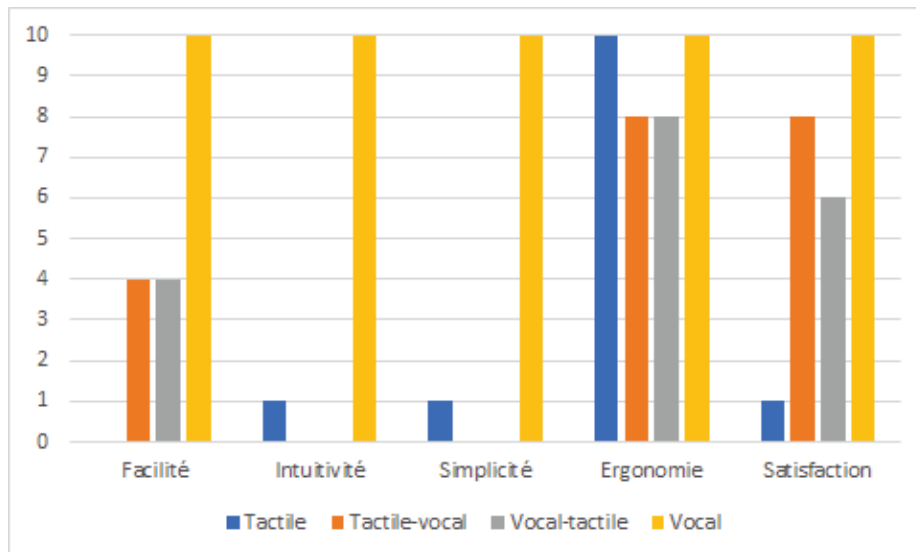


figure 4.22 – Évaluation des quatre tâches d’ajout par l’utilisatrice P3_E2

4.3. RÉSULTATS DE LA DEUXIÈME EXPÉRIMENTATION

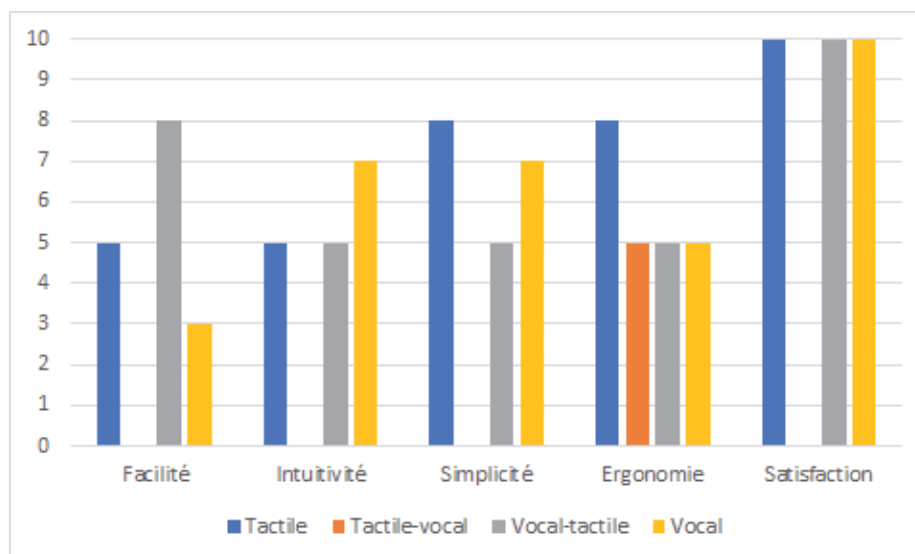


figure 4.23 – Évaluation des quatre tâches d’ajout par l’utilisateur P4_E2

Hors évaluation, les participants ont utilisé le tactile et le vocal à leur guise.

4.3.3 Taux de réussite des actions vocales

Comme lors de la première expérimentation, un certain nombre d’actions vocales sont effectuées. Il est ici montré le nombre d’actions vocales réussies.

Pour chaque participant, le taux d’erreurs est faible. Il se situe entre 5 et 13%. Il faut néanmoins rappeler que l’expérimentation était courte, ce qui fait que l’interaction vocale est peu diversifiée au niveau des interactions avec les fonctionnalités, pour l’expérimentation 1, les utilisateurs ont touché un peu à toutes les fonctionnalités en vocales, tandis que dans l’expérimentation 2, les actions vocales sont centrés sur les rendez-vous.

En observant les figures 4.24, 4.25, 4.26 et 4.27, l’utilisation vocale effectuée par les participants est uniquement centrée sur les rendez-vous et plus particulièrement sur l’ajout de rendez-vous.

Les erreurs vocales sont dues à 50% du temps à une réponse hors sujet de la part de l’utilisateur par rapport à la demande de l’agent virtuel. La reconnaissance vocale et le manque de règles représentent chacun 25% des erreurs vocales restantes.

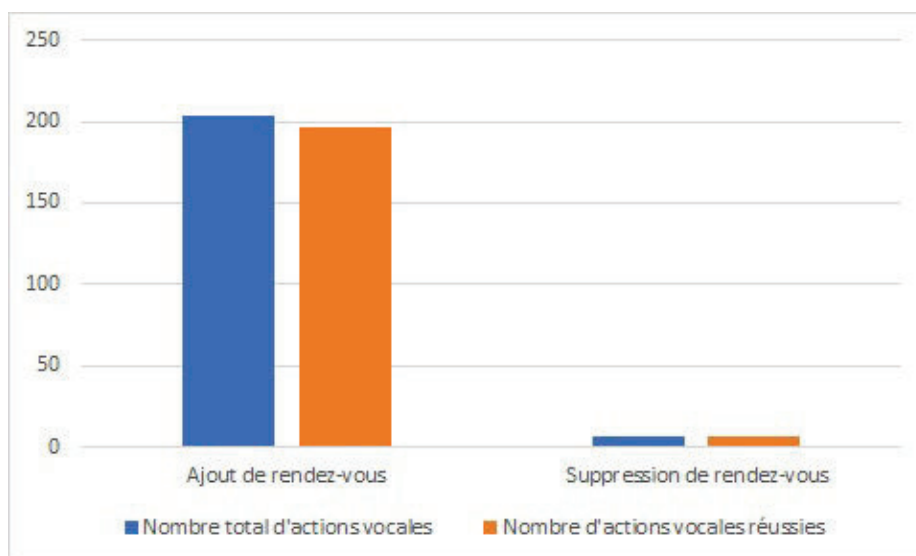


figure 4.24 – Nombre d'actions vocales réussies par la participante P1_E2

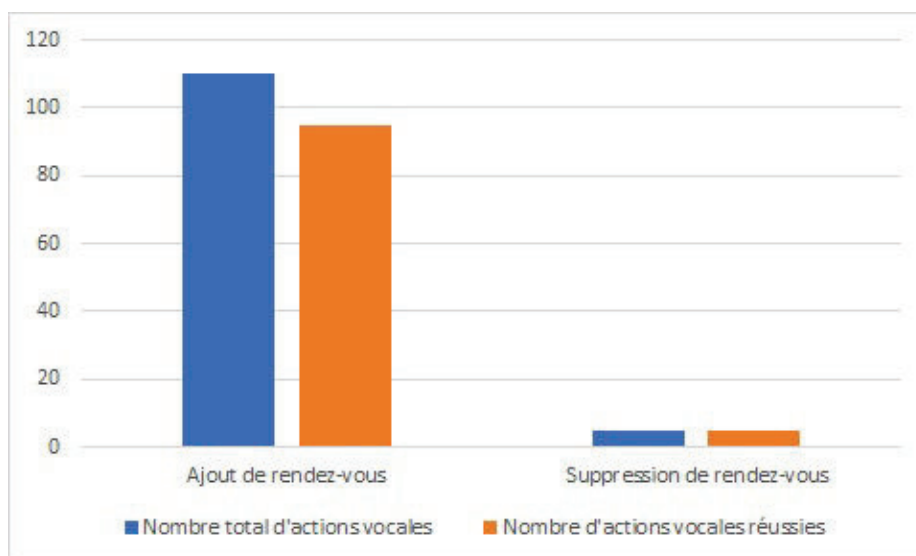


figure 4.25 – Nombre d'actions vocales réussies par la participante P2_E2

4.3. RÉSULTATS DE LA DEUXIÈME EXPÉRIMENTATION

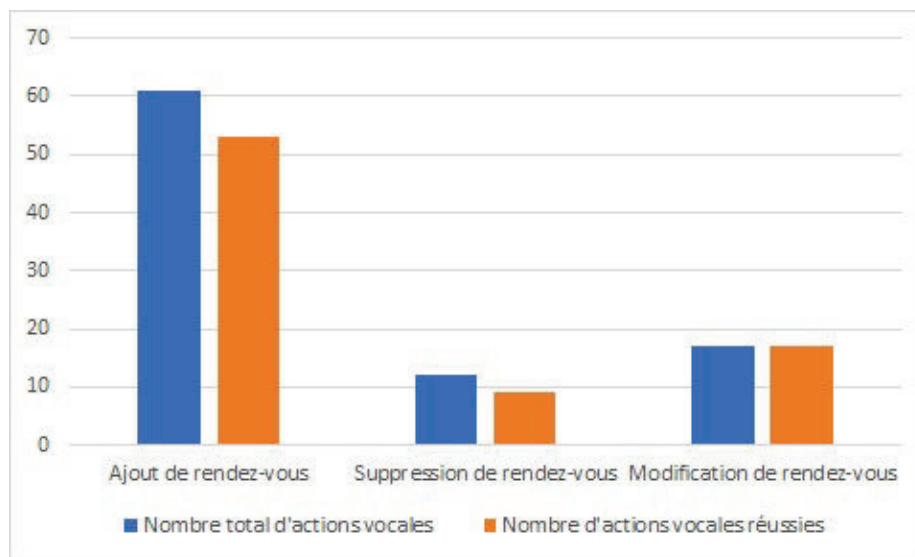


figure 4.26 – Nombre d'actions vocales réussies par la participante P3_E2

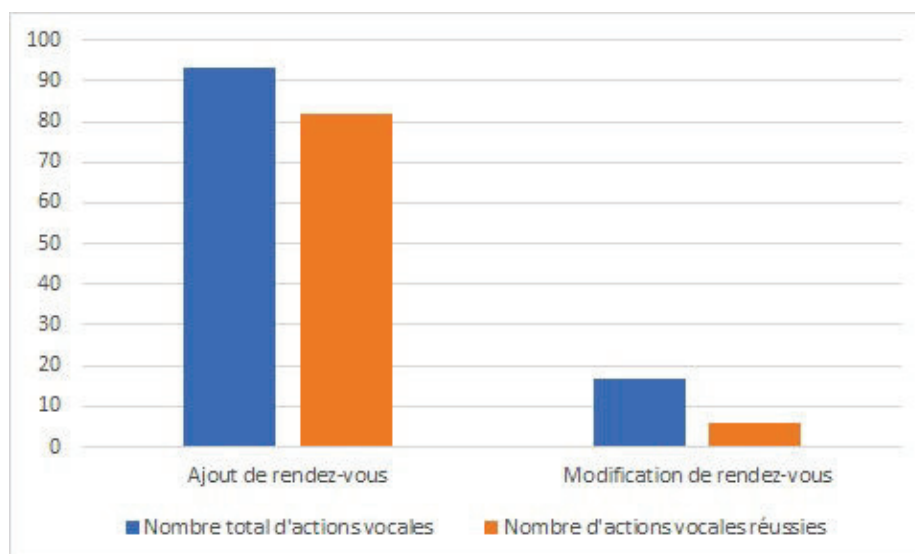


figure 4.27 – Nombre d'actions vocales réussies par le participant P4_E2

Durant les quelques jours d'expérimentation, les utilisateurs ont également utilisé le tactile.

4.3.4 Comparaison de l'utilisation du tactile et du vocal par les participants

Pour chaque figure suivante, il est comparé le nombre total d'actions effectuées en tactile et en vocal selon chaque fonctionnalité. A l'inverse de la première expérimentation, il n'est pas montré l'évolution au cours du temps car l'expérimentation à domicile n'a duré que 4 jours.

Pour la participante P1_E2, la figure 4.28 montre une importante utilisation du tactile pour l'ajout de rendez-vous. Elle a tout de même effectuée quelques interactions tactiles avec d'autres fonctionnalités.

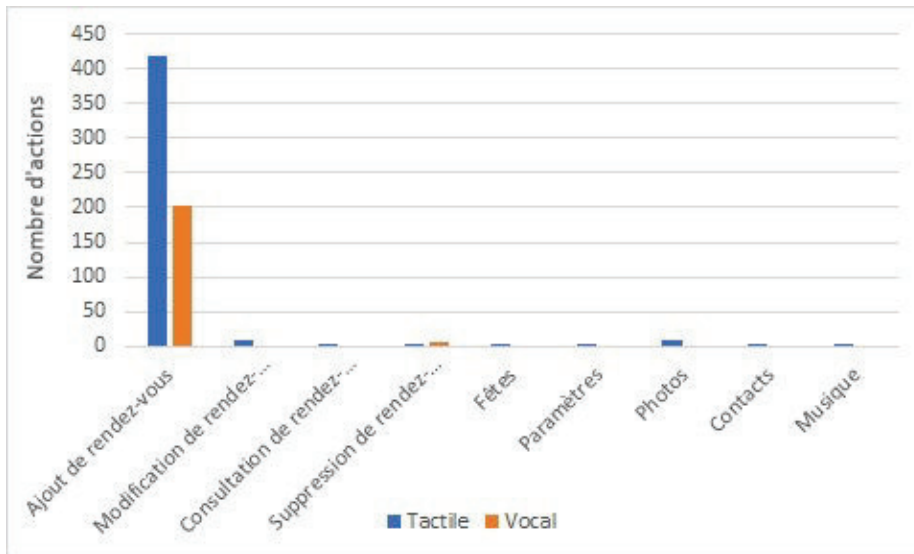


figure 4.28 – Utilisation tactile et vocale de l'application Amelis par la participante P1_E2

De même que la participante P1_E2, la figure 4.29 montre que la participante P2_E2 se sert principalement du tactile pour l'ajout de rendez-vous, mais a également interagi en tactile avec les autres fonctionnalités de l'application. Cette participante ne s'est pas servi de la synchronisation que ce soit du tactile vers le vocal ou du vocal

4.3. RÉSULTATS DE LA DEUXIÈME EXPÉRIMENTATION

vers le tactile.

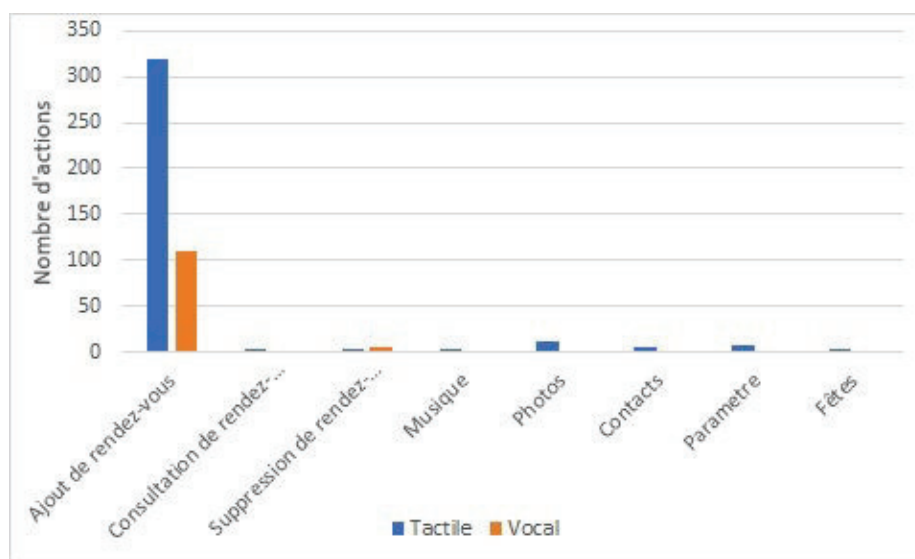


figure 4.29 – Utilisation tactile et vocale de l’application Amelis par la participante P2_E2

Comme le montre la figure 4.30, la participante P3_E2 a principalement utilisé le tactile pour l’ajout de rendez-vous. Cependant, elle a aussi beaucoup interagi en tactile avec la musique, les photos et les paramètres. Cette participante ne s’est pas servi de la synchronisation que ça soit du tactile vers le vocal ou du vocal vers le tactile.

Pour le participant P4_E2, l’utilisation tactile est principalement orientée sur les rendez-vous et plus précisément pour l’ajout et la modification de rendez-vous.

Pour les participants P1_E2 et P4_E2, des synchronisations au niveau de certains ajouts de rendez-vous ont eu lieu. La figure 4.32 reprend les différents cas de synchronisation observés auprès de ces deux participants.

Pour les sous-figures 4.32a et 4.32b, les utilisateurs ont fait face à un conflit horaire après avoir donné une date en vocal. Chacun d’eux à penser à se servir du tactile pour sortir de cette situation. La suite du rendez-vous pouvait ensuite être reprise en vocal ou en tactile. Dans les deux autres cas de synchronisation, 4.32c et 4.32d, le début du rendez-vous à lieu en tactile et les utilisateurs passent ensuite au vocal. Il faut savoir que ces types de synchronisation ont eu lieu après un plusieurs cas de conflits horaires,

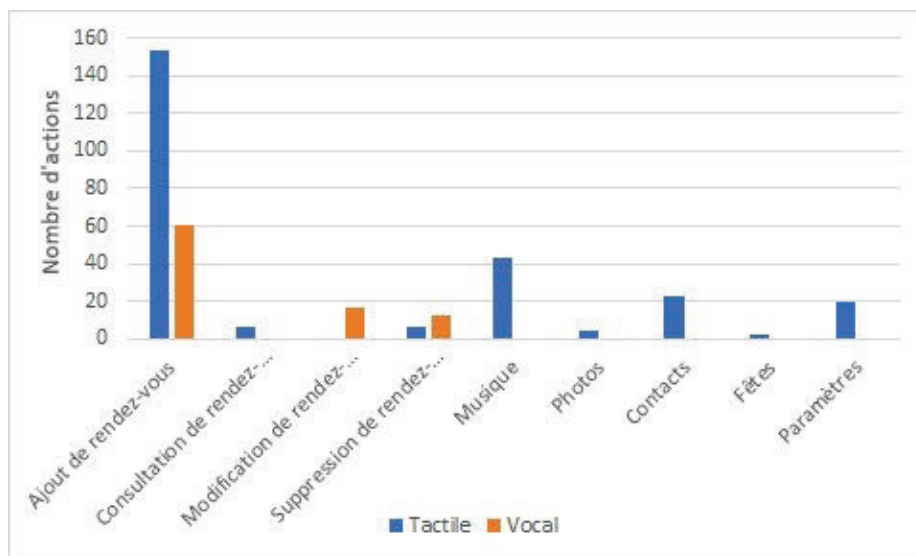


figure 4.30 – Utilisation tactile et vocale de l’application Amelis par la participante P3_E2

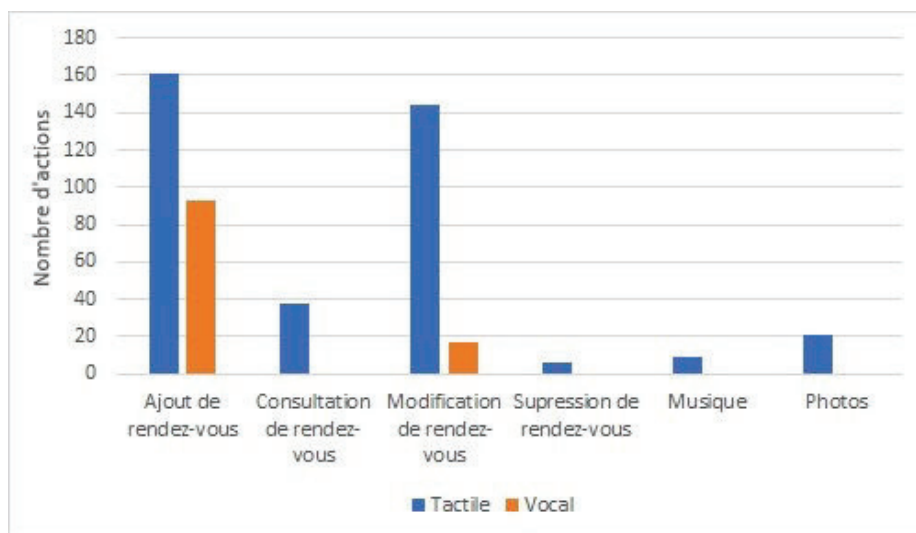
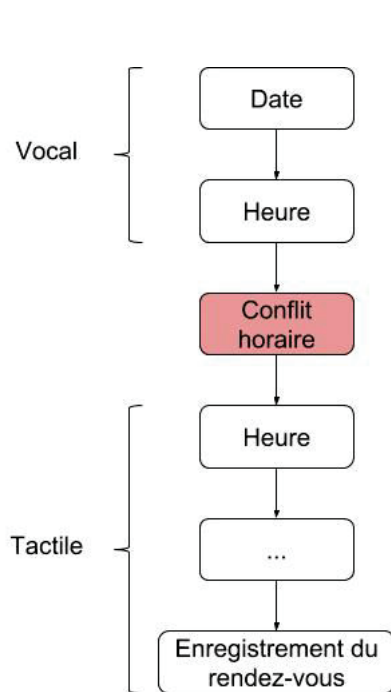
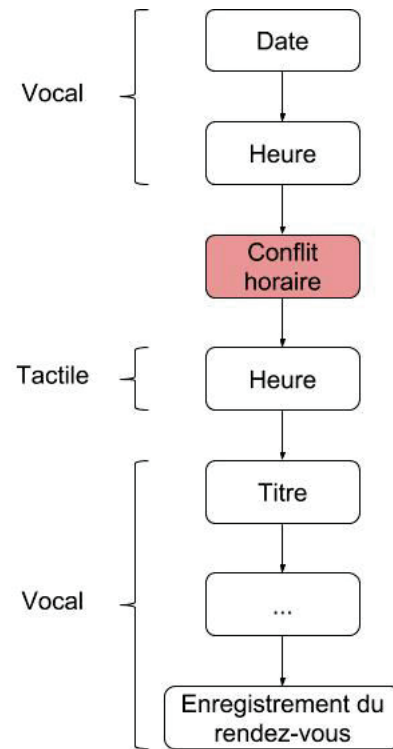


figure 4.31 – Utilisation tactile et vocale de l’application Amelis par la participante P4_E2

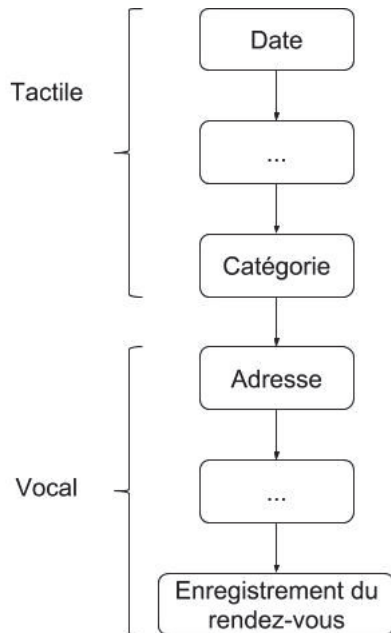
4.3. RÉSULTATS DE LA DEUXIÈME EXPÉRIMENTATION



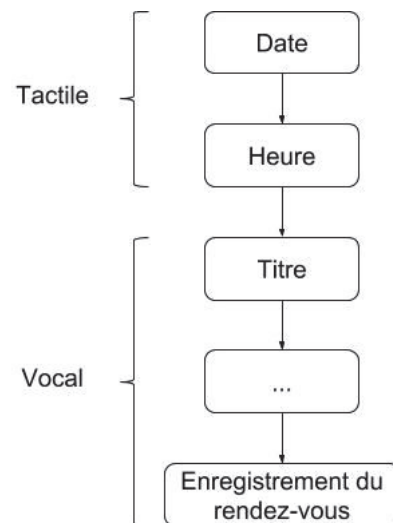
(a) Synchronisation du vocal vers le tactile à cause d'un conflit horaire



(b) Synchronisation du vocal vers le tactile à cause d'un conflit horaire puis retour au vocal



(c) Synchronisation du tactile vers le vocal au milieu d'un ajout de rendez-vous



(d) Synchronisation du tactile vers le vocal à partir du titre du rendez-vous

comme dans les deux cas de synchronisation 4.32a et 4.32b. En plus du dialogue par règles et de la synchronisation, les participants utilisent le dialogue libre.

4.3.5 Utilisation du dialogue libre

Le dialogue libre n'est pas présenté dans les figures précédentes car à l'inverse de la première expérimentation, le dialogue libre et le dialogue par règles sont maintenant deux dialogues distincts. Hormis pour la participante P1__E2, qui n'a pas pu utiliser le dialogue libre à cause d'un problème technique, les dialogues libres entre l'agent virtuel et les autres participants sont présentés ci-dessous.

4.3.5.0.1 Dialogue libre de la participante P2__E2 Utilisateur : parler du film avatar.

Agent virtuel : qu'est ce que tu veux.

Agent virtuel : oh non.

Utilisateur : est ce que tu veux ou non.

Utilisateur : je sais pas quoi quel film.

Agent virtuel : j'essaie.

Agent virtuel : au contraire.

Utilisateur : est ce que tu connais fanny.

Utilisateur : ça veut dire au contraire.

Agent virtuel : juste une seule idée.

Agent virtuel : nous ne l'obtenons pas, on a eu trois de ces amis.

Utilisateur : pas des amis.

Agent virtuel : oui.

Utilisateur : je sais pas quoi dire.

Utilisateur : j'ai vraiment pas d'idée pour ça.

Agent virtuel : ne vous inquiétez pas, nous pouvons trouver le reste de notre temps.

Agent virtuel : juste une douche je veux le voir.

Utilisateur : juste.

Utilisateur : je veux le voir la voiture.

Utilisateur : si tu veux voir.

4.3. RÉSULTATS DE LA DEUXIÈME EXPÉRIMENTATION

Agent virtuel : qu'en est il de.

Agent virtuel : Vouliez-vous voir un rendez-vous, les contacts, les photos ou les fêtes ?

Utilisateur : de ça.

4.3.5.0.2 Dialogue libre de la participante P3_E2 Utilisateur : je prends exposer de parler d'un film d'action avec toi aujourd'hui.

Agent virtuel : Vouliez-vous prends un rendez-vous, une fête ou un contact ?

Utilisateur : non je voulais parler d'un film d'action avec toi aujourd'hui.

Agent virtuel : Avez-vous besoin d'autre chose ?

Utilisateur : parler avec toi d'un film d'action.

Agent virtuel : non je veux sortir d'une liste de fils.

Utilisateur : film.

Agent virtuel : oh oui.

Utilisateur : m'a demandé de te parler de films.

Utilisateur : avec toi.

Agent virtuel : qu'est ce que tu racontes.

Agent virtuel : oui monsieur, je t'aime dans la caméra.

Utilisateur : je t'aime.

Agent virtuel : oui.

Utilisateur : avec toi mais.

Utilisateur : que je vais laisser faire la conversation aujourd'hui.

4.3.5.0.3 Dialogue libre du participant P4_E2 Utilisateur : aujourd'hui on me demande de parler avec l'avatar de films d'action.

Agent virtuel : Vendredi, 22 Septembre, 2017.

Utilisateur : oui.

Utilisateur : je veux parler avec toi dans le film d'action.

Agent virtuel : qu'est ce que tu veux.

Utilisateur : je veux parler avec toi.

Agent virtuel : maintenant je n'ai jamais eu de steak.

Utilisateur : est ce que tu connais le film.

Agent virtuel : j'ai été vu ma première boîte mais je ne peux pas.

Utilisateur : est ce que tu connais le film orange and black.

Agent virtuel : pour un homme.

Utilisateur : non on n'est pas dans la bonne direction est ce que tu connais le film la vengeance de la page.

Agent virtuel : c'est dommage qui soit construit.

Utilisateur : le film était très beau que tu sembles pas l'avoir vu.

Agent virtuel : il fait 9 C et c'est brouillard.

Utilisateur : google.

Agent virtuel : qu'est ce que tu veux.

Utilisateur : je veux te dire c'est terminé on est pas sur la même longueur d'onde.

Agent virtuel : non, je ne sais vraiment pas ce que vous allez faire.

Utilisateur : je vais quitter.

Agent virtuel : je le pense.

Utilisateur : ken meany.

Agent virtuel : je suis désolé tellement je vous ai dit que je vous ai parlé de l'âme de la fucking.

Utilisateur : un très beau langage.

Agent virtuel : la température est de 9 C et c'est brouillard.

Utilisateur : terminé.

Agent virtuel : A plus tard Jean¹.

Avant de discuter des résultats obtenus, les participants ont évalué l'utilisation de l'application Amelis.

4.3.6 Évaluation de l'application Amelis par les participants

Les réponses du tableau 4.5 proviennent de formulaires que les participants ont remplis. À la fin de la semaine d'expérimentation, chaque participant a complété un formulaire par rapport à son ressenti sur l'utilisation de l'application. Avoir une évaluation sur le tactile et le vocal à la fin de l'expérimentation permet d'avoir une

1. Prénom fictif

4.3. RÉSULTATS DE LA DEUXIÈME EXPÉRIMENTATION

comparaison avec l'évaluation des premières rencontres au début de la semaine d'expérimentation, pour voir s'il y a eu une évolution. De plus, cette évaluation permet de cerner plus précisément les problèmes auxquels font face les utilisateurs, et ainsi de pouvoir faire des corrélations avec les résultats obtenus.

tableau 4.5 – Évaluation de l'application Amelis par les participants

		P1	Avis P1	P2	Avis P2	P3	Avis P3	P4
Tactile	Suffisamment de retours sur les actions	Oui		Oui		Non		Non
	Retours sur actions pertinents	Oui		Oui		Non		Oui
	Interaction tactile simple	Oui		Oui		Oui		Non
	Plus simple avec souris/clavier	Non		Non		Non		Oui
	L'interaction tactile répond aux attentes	Oui		Oui		Oui		Oui
Vocal	Suffisamment de retours sur les actions	Oui		Oui		Oui	L'agent virtuel ne comprend pas certaines directives	Non
	Avatar assez rapide pour répondre	Oui		Oui		Oui		Non
	Réponses compréhensibles	Ne sait pas		Oui		Non	Pas toujours rapport	Non
	Savez-vous quand parler	Oui		Oui	Je commence à parler trop vite	Non		Non
	Savez-vous si l'avatar écoute	Oui		Oui		Non		Oui
	Difficultés pour se faire comprendre	Non	Il faut articuler et être clair	Non		Oui	Parfois	Oui
	L'avatar donne-t'il trop d'informations	Non		Non		Non	L'agent virtuel ne comprend pas donc ne m'informe pas	Non
	Sensation d'être maître face à l'avatar	Oui		Oui		Oui		Non
	L'avatar prévient s'il ne comprend pas	Oui		Oui		Non		Oui
	Retours sur actions pertinents	Oui		Oui		Oui		Non
	Interaction vocale simple	Oui		Oui		Non	Non si l'agent ne comprend pas	Non
	Meilleure interaction si l'avatar explique ce qui peut être fait	Non		Oui		Non	L'agent ne saisit pas le sens de certaines questions	Oui
	L'interaction vocale répond aux attentes	Oui		Oui		Oui		Non

Chapitre 5

Discussion

Plusieurs hypothèses sont posées dans le paragraphe 2.3 sur les objectifs au chapitre 2. Pour chaque hypothèse sont associés des résultats qui sont discutés pour expliquer si l'hypothèse est validée.

5.1 Objectif 2 : Offrir un contrôle de l'application Amelis par la parole

Cet objectif contient trois hypothèses.

5.1.1 Contrôle de l'application selon un dialogue par règles

Pour l'hypothèse 1, qui est : "Un dialogue par règles permet d'extraire les informations nécessaires pour contrôler une partie de l'application Amelis et de générer les réponses de l'agent virtuel", le paragraphe 4.1.1 sur les tests techniques E2E montre que le dialogue par règles permet de contrôler une partie de l'application Amelis et de donner des réponses à l'utilisateur. En effet, avec 638 tests, il est possible d'affirmer que le dialogue par règles offre un large choix d'actions réalisables pour contrôler certaines parties de l'application Amelis. De plus, il est vérifié que les réponses de l'agent virtuel sont cohérentes par rapport aux demandes effectuées.

5.1. OBJECTIF 2 : OFFRIR UN CONTRÔLE DE L'APPLICATION AMELIS PAR LA PAROLE

En contrepartie, l'utilisation faite par les participants apporte un point de vue plus mitigé. Il est constaté que les interactions vocales entre l'agent virtuel et l'utilisateur n'aboutissent pas toujours. Selon les figures 4.13, 4.12, 4.14 et 4.15 qui montrent les causes des erreurs d'actions vocales, il apparaît qu'il manque des règles pour contrôler l'application Amelis. En effet, bien qu'il y ait un grand nombre de règles mises en place dans l'expérimentation 1, certaines situations n'ont pu être résolues dû à un manque de règles. Dans la deuxième expérimentation, le pourcentage d'erreurs vocales est beaucoup moins important que lors de la première expérimentation. Cela n'est pas visible dans les figures du chapitre des résultats, mais en comparant les impasses du dialogue par règles entre l'agent virtuel et l'utilisateur des deux expérimentations, il est constaté que les utilisateurs sortent le plus souvent des impasses lors de la deuxième expérimentation, et ce grâce à la synchronisation. De plus, toutes les erreurs provenant du dialogue par règles qui traitait le dialogue libre lors de la première expérimentation, ne sont plus présentes car le dialogue libre est désormais un module à part entière traité par un réseau de neurones. On peut donc affirmer que le dialogue par règles permet de contrôler une partie de l'application Amelis et de fournir des réponses pour l'utilisateur, mais le contrôle ne peut être total que si tous les cas d'utilisation sont traités par les règles ou qu'une synchronisation aide à traiter les impasses du dialogue. Or d'une personne à une autre, de nouveaux cas d'utilisation peuvent apparaître. Cela montre que le contrôle du dialogue par règles comporte également des limites.

5.1.2 Formulations diverses des commandes dans le dialogue par règles

D'après l'hypothèse 2 : L'agent virtuel est capable avec le dialogue par règles de commander l'application Amelis avec diverses formulations. Cette hypothèse est liée à l'hypothèse précédente. En effet, il est montré à travers les tests E2E qu'un grand nombre de formulations est possible. Un exemple est repris dans la figure 4.2, montrant différents exemples de formulations pour l'ajout de rendez-vous. Néanmoins, l'hypothèse est contre balancée à cause des erreurs dues à un manque de règles lors de l'utilisation de l'application Amelis avec les participants lors des premières expé-

rimentations. Pour que l’agent virtuel soit capable, à l’aide du dialogue par règles, de commander l’application Amelis avec diverses formulations, l’ensemble de règles doit être suffisamment large. Par exemple, le fait que les utilisateurs aient pris des rendez-vous la nuit durant la première expérimentation est dû à un manque initial au niveau des règles, il a fallu demander aux utilisateurs si l’heure qu’ils fournissaient correspondait au matin ou à l’après-midi. Dans la conception ont été évoqués différents moyens pour élargir les règles : intégration de synonymes et combinaisons de règles. Cependant, lors des tests avec les utilisateurs, ceux-ci s’expriment avec de nouvelles manières qui n’avaient pas été initialement prévues.

Pour contrôler pleinement l’application Amelis avec différentes formulations, il faut récolter autant d’exemples que possible pour la formulation de chaque tâche. Cela peut se faire de façon manuelle en trouvant de nouveaux exemples et les ajouter aux règles, ou de façon automatique en concevant un module capable de trouver sur internet des formulations similaires à la demande. Par exemple, pour demander la météo, il est possible de demander à des personnes d’écrire toutes les formulations auxquelles elles pensent pour demander la météo. Une autre solution serait de composer non pas un dictionnaire de synonymes mais un dictionnaire de phrases similaires. Dans l’article de Han J. et al. [102], il est montré qu’à partir d’un large corpus de questions, il est possible de trouver les questions similaires à une nouvelle question. Cela nécessite d’entraîner un modèle pour qu’il soit capable de retrouver des phrases similaires. Par ces biais, il serait alors possible d’élargir les règles pour prendre en compte le plus de formulations possible. Plus le nombre d’exemples de formulations diverses regroupées est grand, plus les règles peuvent s’agrandir. Cependant, il est toujours possible que des exceptions surviennent et selon le public visé, certains cas peuvent être omis. Par exemple, un public de plus de 60 ans et plus n’utilisera pas forcément le même langage qu’un public adolescent.

5.1.3 Demandes intelligibles de l’agent virtuel

L’hypothèse 3 suppose que l’utilisateur est capable de comprendre les demandes de l’agent virtuel pour y répondre. Les figures 4.13, 4.12, 4.14 et 4.15 de la première expérimentation permettent de répondre à l’hypothèse. Une cause des erreurs vocales

5.2. OBJECTIF 3 : OFFRIR DES CONTEXTES DE DISCUSSION LIBRE ENTRE L'AGENT VIRTUEL ET L'UTILISATEUR POUR DONNER UN ASPECT PLUS HUMAIN LORS DES ÉCHANGES

est une réponse inappropriée de l'utilisateur à une demande de l'agent virtuel. Lors de la première expérimentation, hormis la participante P2_E1 qui a peu utilisé le vocal, les autres participants ont utilisé le vocal entre une cinquantaine et une cent vingtaines de fois au cours de ces trois mois. Dans la majorité des cas, l'utilisateur est capable de comprendre les demandes de l'agent virtuel et d'y répondre, le plus haut pourcentage d'erreurs vocales dû à des réponses hors sujet durant la première expérimentation est de 14%. Il est noté que les erreurs vocales liées à des réponses hors sujet surviennent lors des ajouts de rendez-vous, 2 fois pour la participante P1_E1 et 7 fois pour le participant P4_E1. Cela est probablement dû au fait que l'ajout de rendez-vous est une tâche plus complexe et demande une interaction plus longue. Puisque l'ajout de rendez-vous contient plusieurs champs à compléter et que l'utilisateur connaît les informations qu'il souhaite ajouter, ce dernier a tendance à donner ces informations, sans prêter attention aux demandes de l'agent virtuel, en pensant que l'agent virtuel saura comprendre les informations reçues. Cela montre d'une part que l'utilisateur peut comprendre et répondre à l'agent virtuel, mais d'autre part qu'il est nécessaire que l'agent virtuel puisse s'adapter aux réponses de l'utilisateur. Ce point nous ramène à la discussion sur l'hypothèse précédente où l'agent virtuel doit être capable de comprendre plusieurs types de formulations.

5.2 Objectif 3 : Offrir des contextes de discussion libre entre l'agent virtuel et l'utilisateur pour donner un aspect plus humain lors des échanges

Cet objectif contient deux hypothèses.

5.2.1 Réseaux de neurones et dialogue libre

Selon l'hypothèse 1 : Le dialogue libre nécessite d'utiliser les réseaux de neurones pour le modéliser

Dans le paragraphe 5.1 précédent sur le dialogue par règles, il est montré que les règles sont limitées dès lors qu'une nouvelle exception survient. Il faut prendre en

compte que ces résultats concernaient l'utilisation du dialogue par règles pour des fonctionnalités déjà prévues. Par exemple, la météo est une fonctionnalité accessible en vocal, il faut donc instaurer toutes les règles autour pour permettre d'interagir avec. En se concentrant simplement sur les erreurs vocales dues à un manque de règles au niveau du dialogue par règles durant la première expérimentation, il est constaté que ces erreurs représentent :

- 25% pour la participante P1_E1 ;
- 92% pour la participante P2_E1 ;
- 50% pour la participante P3_E1 ;
- 90% pour le participant P4_E1.

Un grand nombre de tentatives de discussion libre échoue à cause d'un manque de règles pour pouvoir traiter chaque cas spécifique. Pour répondre à cette demande de discussion de la part de l'utilisateur, il faut construire un module de discussion libre qui peut couvrir n'importe quel sujet de conversation. Ceci ne peut pas être réalisé avec des règles qui couvriraient tous les sujets possibles. Toutefois, la demande de la part des utilisateurs pour avoir une discussion libre amène d'autres algorithmes de traitement de langage naturel. Comme vu lors de la revue de littérature, les réseaux de neurones permettent de construire automatiquement des réponses.

5.2.2 Qualité du dialogue libre et acceptabilité du dialogue de l'agent virtuel

L'hypothèse 2 stipule que : La qualité du dialogue libre impacte sur l'acceptabilité du dialogue avec l'agent virtuel. Il est prouvé dans l'hypothèse précédente que la qualité du dialogue libre n'est pas suffisante lorsque le dialogue par règles traite le dialogue libre. La seconde expérimentation vient appuyer l'hypothèse 2. En effet, elle montre ce qu'apporte la mise en place d'un dialogue libre à l'aide de réseaux de neurones. En comparant les discussions libres des participants avec l'agent virtuel présentées dans le chapitre précédent sur les résultats, la conversation entre l'utilisateur et l'agent virtuel tourne rapidement en un dialogue peu compréhensible avec l'agent virtuel. Les utilisateurs tentent au départ de se reprendre ou d'essayer d'expliquer à l'agent virtuel ce dont ils veulent lui parler, mais ils constatent que la conversation

5.2. OBJECTIF 3 : OFFRIR DES CONTEXTES DE DISCUSSION LIBRE ENTRE L'AGENT VIRTUEL ET L'UTILISATEUR POUR DONNER UN ASPECT PLUS HUMAIN LORS DES ÉCHANGES

reste sans queue ni tête. Les utilisateurs finissent par abandonner la discussion en gardant en tête que l'agent virtuel ne les comprend pas lorsqu'ils discutent avec.

Pour la première expérimentation, les tableaux 4.1, 4.2, 4.3, 4.4 montre une diminution de l'utilisation du vocal. Les discussions entre les utilisateurs et l'agent virtuel ne montrent pas une cause spécifique à cette diminution. L'utilisation du dialogue libre est répartie sur le temps. De plus bien que les participants font face à plusieurs erreurs lors de l'utilisation du vocal, ils continuent à utiliser le dialogue par la suite. Nous supposons que la diminution de l'utilisation du vocal s'est faite progressivement pour trois raisons :

- Les participants ont appris à se servir du tactile, le résultat n'aboutissait pas à un échec et ils avaient accès à plus de fonctionnalités ;
- Le dernier mois d'expérimentation présente une diminution de l'utilisation de l'application, que ça soit en vocal ou en tactile ;
- Les utilisateurs ont manifesté avoir moins d'intérêts à entrer des rendez-vous le dernier mois car ils ne pourraient pas se les faire rappeler, puisque le calendrier allait leur être ôté.

Pour la deuxième expérimentation, l'utilisation de l'application et des différents dialogues n'est pas suffisante pour conclure sur l'impact que le dialogue libre peut avoir sur le dialogue avec l'agent virtuel. Il est néanmoins constaté que les utilisateurs cessent la discussion avec l'agent virtuel ou s'expriment d'une manière montrant que pour eux l'agent virtuel ne comprend pas. Les utilisateurs sont donc conscients d'un manque de qualité au niveau du dialogue libre. L'apprentissage du réseau de neurones n'est donc pas suffisamment poussé pour obtenir des phrases fluides et sensées. Cependant, l'apprentissage est allé assez loin pour permettre d'avoir une bonne syntaxe. En entraînant plus longuement le réseau de neurones, il serait possible de pallier à l'aspect sémantique qui n'est pas suffisamment présent pour le moment. De plus, le pourcentage d'erreurs vocales pour le dialogue durant la deuxième expérimentation est descendu entre 5 et 13% tandis que le pourcentage était entre 38 et 46% pour la première expérimentation. La qualité du dialogue par règles s'est donc amélioré et impacte sur l'acceptabilité du dialogue avec l'agent virtuel.

5.3 Objectif 4 : Intégrer un mode d'interaction tactile et un mode d'interaction vocal pour offrir un choix d'interaction plus large et plus simple à l'utilisateur

Cet objectif contient deux hypothèses.

5.3.1 Deux modes d'interaction élargissent les possibilités de choix d'interaction

L'hypothèse 1 énonce que : deux modes d'interaction élargissent les possibilités de choix d'interaction. Toutes les fonctionnalités de l'application Amelis sont toutes accessibles en tactile, mais certaines ne sont pas accessibles en vocal, comme le choix d'une musique ou l'ajout d'un contact. Les figures 4.17, 4.16, 4.18 et 4.19 montrent l'utilisation, au cours des semaines, des fonctionnalités accessibles à la fois en tactile et en vocal. Le vocal est utilisé principalement par les participants P1_E1, P3_E1 et P4_E1. La participante P3_E1 a utilisé le vocal 27 fois pour l'ajout de rendez-vous et 26 fois pour accéder aux autres fonctionnalités de l'application Amelis. Les participants P1_E1 et P4_E1 ont surtout utilisé le vocal pour commander des actions sur les rendez-vous, 96 fois pour la participante P1_E1 et 106 fois pour le participant P4_E1. Pour les autres tâches, la participante P1_E1 s'est servie 56 fois du vocal et le participant P4_E1 s'est servi 9 fois du vocal pour les autres fonctionnalités. En comparant l'utilisation de chaque participant, l'accès aux différents menus, à la fois accessible en tactile et en vocal, de l'application Amelis est effectué en majorité en tactile :

- La participante P1_E1 s'est servie 449 fois du tactile pour accéder aux fonctionnalités autres que celles de l'agenda, contre 56 fois en vocal ;
- La participante P2_E1 s'est servie 338 fois du tactile pour accéder aux fonctionnalités autres que celles de l'agenda, contre 13 fois en vocal ;
- La participante P3_E1 s'est servie 462 fois du tactile pour accéder aux fonctionnalités autres que celles de l'agenda, contre 26 fois en vocal ;

5.3. OBJECTIF 4 : INTÉGRER UN MODE D'INTERACTION TACTILE ET UN MODE D'INTERACTION VOCAL POUR OFFRIR UN CHOIX D'INTERACTION PLUS LARGE ET PLUS SIMPLE À L'UTILISATEUR

- Le participant P4_E1 s'est servi 1139 fois du tactile pour accéder aux fonctionnalités autres que celles de l'agenda, contre 9 fois en vocal ;

L'interaction tactile est plus simple et plus rapide lorsque la fonctionnalité ne nécessite que d'un ou deux appuis pour l'activer. Le fait que le vocal soit utilisé pour les rendez-vous peut s'expliquer par le fait que l'interaction peut être plus intuitive et plus courte. En effet, il y a beaucoup plus d'interactions tactiles à réaliser lorsqu'on ajoute un rendez-vous, notamment lorsqu'il faut entrer sur le clavier le titre du rendez-vous, alors qu'en vocal, il suffit de donner la commande et l'agent virtuel peut effectuer l'action à la place de l'utilisateur. Les deux modes d'interaction donnent la liberté à l'utilisateur de choisir le mode qui lui convient le mieux selon l'action qu'il souhaite réaliser.

5.3.2 Synchronisation entre les deux modes d'interactions

D'après l'hypothèse 2 : La synchronisation entre les deux modes d'interactions simplifie leur cohabitation. Dans la deuxième expérimentation, les rencontres avec les utilisateurs montrent qu'ils ne sont pas à l'aise avec la synchronisation à part la participante P2_E2, comme le montre les figures 4.21, 4.22, 4.20 et 4.23. Cependant, les utilisateurs P1_E2 et P4_E2 ont su se sortir d'impasses lorsqu'ils étaient seuls face à l'application Amelis, en changeant de mode d'interaction (voir figure 4.32). La participante P1_E2 et le participant P4_E2 ont solutionné 2 fois une impasse grâce la synchronisation. Outre la synchronisation en cas d'impasse, les participants utilisent aussi la synchronisation pour compléter une action avec l'autre mode plus adapté. Par exemple, la participante P1_E2 complète un ajout d'un rendez-vous en vocal alors qu'elle ne faisait pas face à une impasse en tactile ; les champs comme l'adresse étaient plus simples à remplir en donnant l'information à l'agent virtuel. La synchronisation permet aux deux modes d'interaction de recevoir les mêmes informations et l'utilisateur peut ainsi changer de mode selon ses besoins tout en évitant d'avoir à tout recommencer dans le nouveau mode d'interaction choisi. Ainsi, tant que la synchronisation n'était pas implantée, comme lors de la première expérimentation, si l'utilisateur commençait un ajout de rendez-vous en tactile et qu'il voulait le continuer en vocal, il devait redonner toutes les informations en vocal.

5.4 Objectif 5 : Évaluer les performances techniques et l'impact sur l'utilisateur du dialogue de contrôle et libre

5.4.1 La reconnaissance vocale comprend l'utilisateur

L'hypothèse 1 stipule que : l'utilisateur est correctement compris par le modèle de reconnaissance vocale. Le module de reconnaissance vocale de Google a été choisi car il est performant, rapide, et comprend le français. Dans la présentation des participants dans le chapitre 4 sur les résultats, le taux de reconnaissance vocale pour la lecture des textes sans pause varient entre 15 et 75% de réussite. L'exercice étant compliqué, nous nous sommes basés sur le taux de reconnaissance vocale pour les textes avec des pauses entre chaque phrase, car cela est plus représentatif sur la manière dont une personne parle. Pour cet exercice, le pourcentage de reconnaissance se situe à environ 75% de réussite pour chaque participant. En regardant les figures 4.17, 4.16, 4.18 et 4.19 de la première expérimentation ou les discussions entre l'agent virtuel et l'utilisateur de la deuxième expérimentation, des problèmes de reconnaissance vocale reviennent souvent. Les problèmes de reconnaissance vocale des deux expérimentations oscillent entre 20 et 30% des erreurs vocales. Diverses causes expliquent les erreurs de reconnaissance vocale :

- L'articulation. Dans le tableau 4.5 sur les retours des utilisateurs de la deuxième expérimentation, la participante P1_E1 constate que si elle prend le temps d'articuler en prononçant ses mots, elle est plus facilement reconnue ;
- La vitesse de prononciation. Plus l'utilisateur parle vite, moins il articule, ce qui complique la reconnaissance ;
- L'accent de l'utilisateur. Certains participants ont un accent prononcé, que ce soit un accent québécois ou espagnol comme la participant P1_E1. Le système de reconnaissance de Google est entraîné pour reconnaître des mots français avec un accent de France. Même si le critère d'inclusion des participants nécessitait une reconnaissance vocale supérieure à 70%, certaines prononciations québécoises peuvent être plus difficiles à comprendre ;

5.4. OBJECTIF 5 : ÉVALUER LES PERFORMANCES TECHNIQUES ET L'IMPACT SUR L'UTILISATEUR DU DIALOGUE DE CONTRÔLE ET LIBRE

- L'apprentissage du système est influencé par l'apprentissage et par les mots appris. Par exemple, l'utilisateur P4_E1 a dit "10 heures" et le système de reconnaissance a reconnu le terme "Deezer".

Un système de reconnaissance vocale entraîné pour l'utilisateur d'Amelis pourrait améliorer les résultats. Si le public visé a des caractéristiques spécifiques, telles que l'accent, ou que le projet pour lequel la reconnaissance vocale est utilisé touche un milieu plus particulier, par exemple la prise de rendez-vous, il peut être intéressant d'entraîner soi-même un système de reconnaissance vocale [103]. Cela permet de s'adapter à l'utilisateur et aussi au domaine visé. La confusion entre 10 heures et Deezer serait ainsi évitée.

5.4.2 Validation par les tests techniques

Suivant l'hypothèse 2 : des tests techniques permettent de valider le bon fonctionnement de l'application Amelis. Trois types de tests techniques sont effectués, les tests E2E pour le dialogue par règles, et les tests BLEU et sémantiques pour le réseau de neurones. Les tests E2E permettent d'identifier les erreurs dans l'application et ainsi de refactoriser le code pour s'approcher le plus possible des attentes de l'utilisateur. Ils permettent à la fois de vérifier le bon fonctionnement du tactile et du vocal. En outre, ils valident que les informations échangées lors d'une synchronisation sont justes et que les deux modes d'interaction reçoivent les mêmes informations. De plus, ils testent un grand nombre de formulations de demandes de la part de l'utilisateur pour vérifier que les règles couvrent suffisamment de cas. Les 638 tests E2E permettent ainsi de valider les fonctionnalités actuellement mises en place, ainsi que les fonctionnalités manquantes. En effet, des tests sont réalisés sur des cas non traités pour vérifier que l'agent virtuel aide l'utilisateur. Cela veut dire qu'au lieu d'agrandir indéfiniment les règles, il est possible de concentrer les règles sur les fonctionnalités les plus utilisées en vocal. Pour des fonctionnalités moins utilisées de manière vocale il est possible de créer une règle commune qui explique à l'utilisateur que telle fonctionnalité n'est pas prise en charge par l'agent virtuel et d'inciter l'utilisateur à se servir du tactile.

Les tests techniques concernant l'apprentissage du réseau de neurones valident les résultats lors de l'utilisation du dialogue libre. En effet, le test sémantique donne un

pourcentage de similarité sémantique de 9% entre la réponse attendue et la réponse obtenue. Cela est vérifié par les extraits de dialogue libre de la seconde expérimentation, où les réponses de l'agent virtuel présentent une sémantique mal construite. Le test BLEU, dont la valeur est de 2, montre que les réponses obtenues ont peu de sens et ce résultat est confirmé par les dialogues libres de la deuxième expérimentation. Les réseaux de neurones actuels utilisant des modèles "sequence to sequence" atteignent un score compris entre 25 et 40 [75], [97]. Ces modèles permettent d'obtenir des réponses compréhensibles. Néanmoins, bien que l'apprentissage ne soit pas suffisamment développé, il arrive que la syntaxe soit bonne. Un autre point à noter est que lors du dialogue libre, le dialogue par règles prend le dessus. Cela est désiré puisque nous avons considéré l'exécution d'une règle plus importante que le dialogue libre. Par exemple, dans le dialogue libre de la participante P2_E2, la participante dit "Je veux le voir la voiture. Si tu veux voir". L'agent virtuel répond alors "Voulez-vous voir un rendez-vous...". Ceci est dû à l'activation d'une règle.

5.4.3 Appréciation du dialogue avec l'agent virtuel par les utilisateurs

L'hypothèse 3 dit que : Les utilisateurs apprécient le dialogue avec l'agent virtuel et trouvent plus facile l'utilisation de l'application Amelis avec l'agent virtuel. En regardant les tableaux 4.2, 4.1, 4.3 et 4.4, l'utilisation vocale est moins fréquente que le tactile. Une première raison est que les fonctionnalités ne sont pas toutes accessibles en vocal. Il est constaté que tout au long des expérimentations 1, l'utilisation du vocal diminue, voire cesse d'exister, alors que l'utilisation du tactile subit une faible baisse d'utilisation en comparaison. L'expérimentation 2 est trop courte pour affirmer que le tactile ou le vocal est privilégié. Cependant, la figure 4.20 montre que l'utilisatrice P1_E1, qui n'était pas à l'aise avec la synchronisation du vocal vers le tactile, l'a pourtant utilisé 4 fois pour finir l'ajout de rendez-vous. Durant les expérimentations 1 et 2, certaines réponses montrent une incompréhension de la part des utilisateurs face à l'agent virtuel. Par exemple, le participant P4_E2 a répondu lors d'une discussion libre "Je veux te dire c'est terminé, on est pas sur la même longueur d'onde". En discutant avec les utilisateurs suite aux expérimentations, ils disent avoir l'impression

5.4. OBJECTIF 5 : ÉVALUER LES PERFORMANCES TECHNIQUES ET L'IMPACT SUR L'UTILISATEUR DU DIALOGUE DE CONTRÔLE ET LIBRE

de ne pas être compris par l'agent virtuel. Certains ont dit qu'ils apprécieraient discuter avec l'agent virtuel si ce dernier était capable de tenir une conversation normale et de comprendre leurs demandes, quelles soient complexes ou non. Les utilisateurs préfèrent se servir du tactile pour interagir avec l'application Amelis, mais sont prêts à plus utiliser le vocal si ce dernier est plus fonctionnel.

Conclusion

Ce mémoire a présenté le contexte de recherche et la problématique de la conception d'un dialogue entre les aînés et un agent virtuel au sein d'une application. Ce mémoire a réussi à faire ressortir les difficultés liées à la mise en place d'un tel dialogue, ainsi que les besoins, les habitudes et l'adaptation des utilisateurs face au dialogue instauré dans l'application. L'intégration d'une modalité vocale dans un calendrier tactile a fait l'objet de conception informatique de plusieurs ordres :

- L'architecture logicielle du calendrier a dû être étendue avec une modalité vocale synchronisée sur les interactions tactiles ;
- Un dialogue par règles a été conçu et permet d'utiliser des synonymes ;
- Un dialogue libre basé sur l'apprentissage de réseaux de neurones a été intégré pour étendre le dialogue et ne pas le limiter qu'à des commandes pour l'application Amelis ;
- Un passage entre les deux modes de dialogue (dialogue par règles et dialogue libre) a été conçu pour que les commandes du calendrier soient prioritaires en tout temps.

Les deux expérimentations menées auprès de 8 personnes ont permis d'apporter un contraste sur les tests techniques de l'application. Elles ont permis d'observer l'utilisation réelle de l'application par des personnes âgées. Cela a mis en avant les manques et les points forts du dialogue homme-machine auprès d'aînés.

Le contrôle d'une application est possible en vocal si le dialogue mis en place permet de comprendre les différentes formulations de l'utilisateur. Bien que l'utilisateur puisse adapter son dialogue pour se faire comprendre par un agent virtuel, il faut mettre en place un dialogue capable de comprendre l'utilisateur quelque soit sa manière de s'exprimer. Cela permet à l'utilisateur d'être en confiance avec l'utilisation

CONCLUSION

du vocal, ainsi que de lui permettre de déléguer des tâches plus complexes en tactile.

Pour que le dialogue soit complet, il faut que l’agent virtuel soit capable de discuter avec l’utilisateur de manière libre tout en utilisant des phrases structurées et sensées. Cela implique la mise en place d’un dialogue conçu pour répondre à n’importe quel sujet de discussion.

L’apport d’une interaction vocale en plus d’une interaction tactile permet à l’utilisateur d’adapter ses interactions selon ses propres besoins pour tirer le meilleur retour possible sur l’utilisation de l’application. De plus, le lien entre les deux modes d’interactions donne la possibilité à l’utilisateur de changer de mode selon les tâches tout en gardant l’information à jour entre ces deux modes.

Fournir un dialogue entre une personne et un agent virtuel dans le cadre de la commande d’un calendrier est donc possible. Cependant, certaines difficultés amenant à des améliorations futures sont à souligner :

- Le dialogue par règles permet de contrôler une application, mais il est nécessaire d’identifier tous les cas d’utilisation possible et de pouvoir reconnaître des données difficiles à récupérer ;
- L’utilisation d’un réseau de neurones pour créer un dialogue libre est avantageuse, mais l’apprentissage doit être suffisamment pointu pour obtenir des réponses structurées et sensées. Cela implique d’avoir un réseau de neurones avec des paramètres adaptés et d’avoir un corpus de texte suffisamment large pour permettre à l’apprentissage de couvrir le plus de sujets possible ;
- La reconnaissance vocale nécessite un apprentissage spécifique pour s’adapter à chaque utilisateur, que ce soit son accent, sa prononciation ou sa rapidité d’élocution.

Limiter la fracture numérique des personnes âgées requiert d’augmenter la convivialité des interfaces. Proposer une modalité vocale avec un agent virtuel est une piste que les participants ont appréciée. Toutefois, la réalisation d’un agent virtuel et les expérimentations avec les aînés indiquent qu’il est indispensable de s’assurer de la qualité du dialogue, que la modélisation du dialogue par règles peut faciliter l’utilisation d’applications où le tactile demande plus d’entrées de clavier, et que le dialogue doit être intégré à une utilisation tactile avec une synchronisation.

Mes contributions propres à ce projet de recherche ont été :

CONCLUSION

- De mettre en place l’architecture tactile suite à des rencontres avec les aînés pour respecter leurs besoins et leur attentes sur les fonctionnalités à intégrer au calendrier ;
- De concevoir le dialogue par règles et de le développer. Cela comprends la refactorisation des règles et le développement de synonymes, dans le but d’améliorer les interactions vocales entre l’agent virtuel et les aînés, et de répondre au mieux à leur attentes ;
- De développer un module de dialogue libre pour permettre de répondre au besoin des aînés de discuter de sujets divers avec l’agent virtuel ;
- De mettre en place un système de communication entre une modalité vocale et une modalité tactile, permettant à la fois de fournir suffisamment de retours vocaux et visuels propres pour les aînés, et à la fois d’offrir plus de choix d’interaction et d’adaptabilité pour les aînés ;

Les perspectives de cette recherche sont de fournir un dialogue avec un meilleur taux de reconnaissance pour commander le calendrier en étendant les formulations possibles. La recherche du dialogue libre doit être poursuivie pour offrir aux personnes âgées la convivialité de dialoguer avec l’agent virtuel sans que la conversation ne devienne incompréhensible.

Annexe A

Maquettes du calendrier Amelis

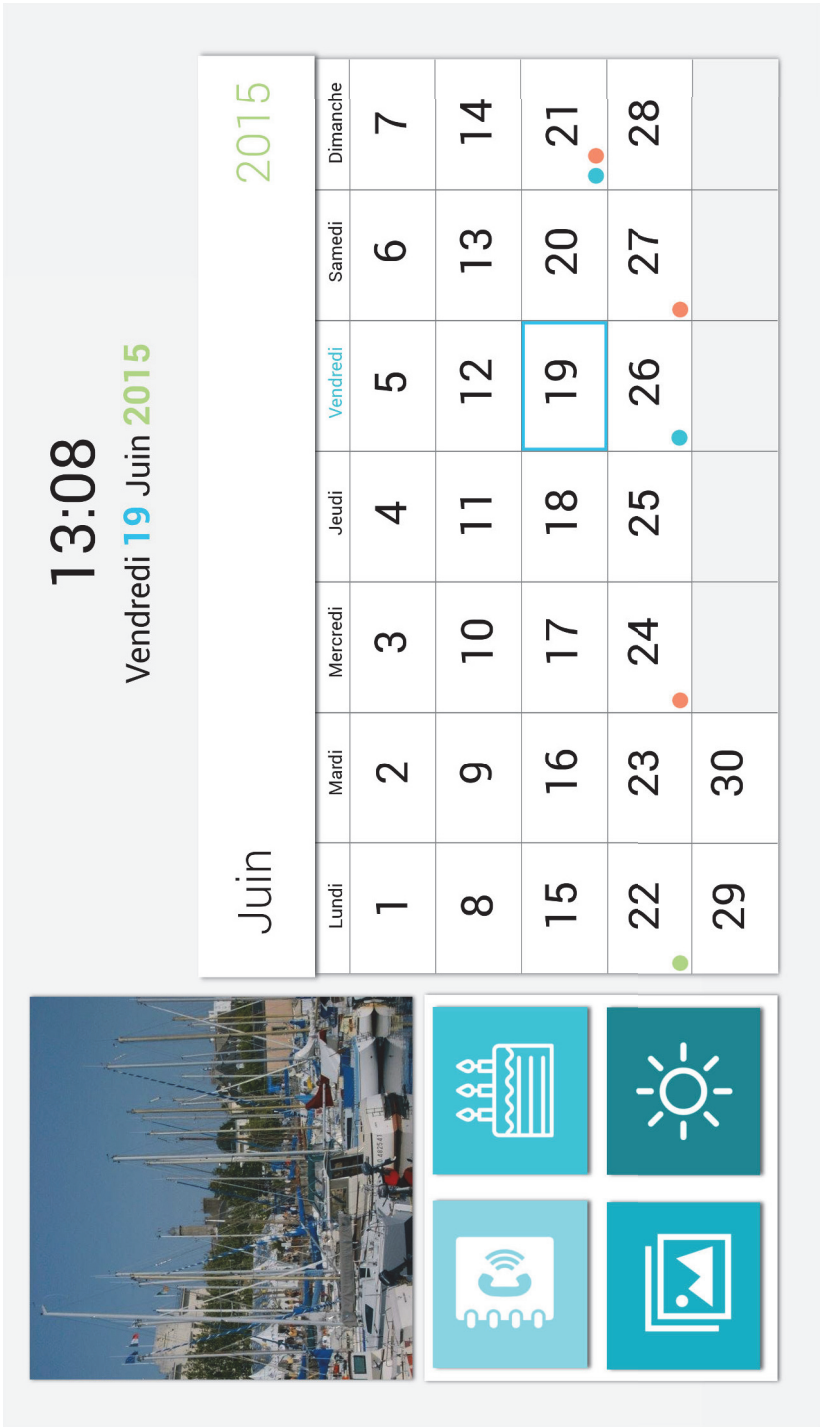


figure A.1 – Maquette du calendrier Amelis en France

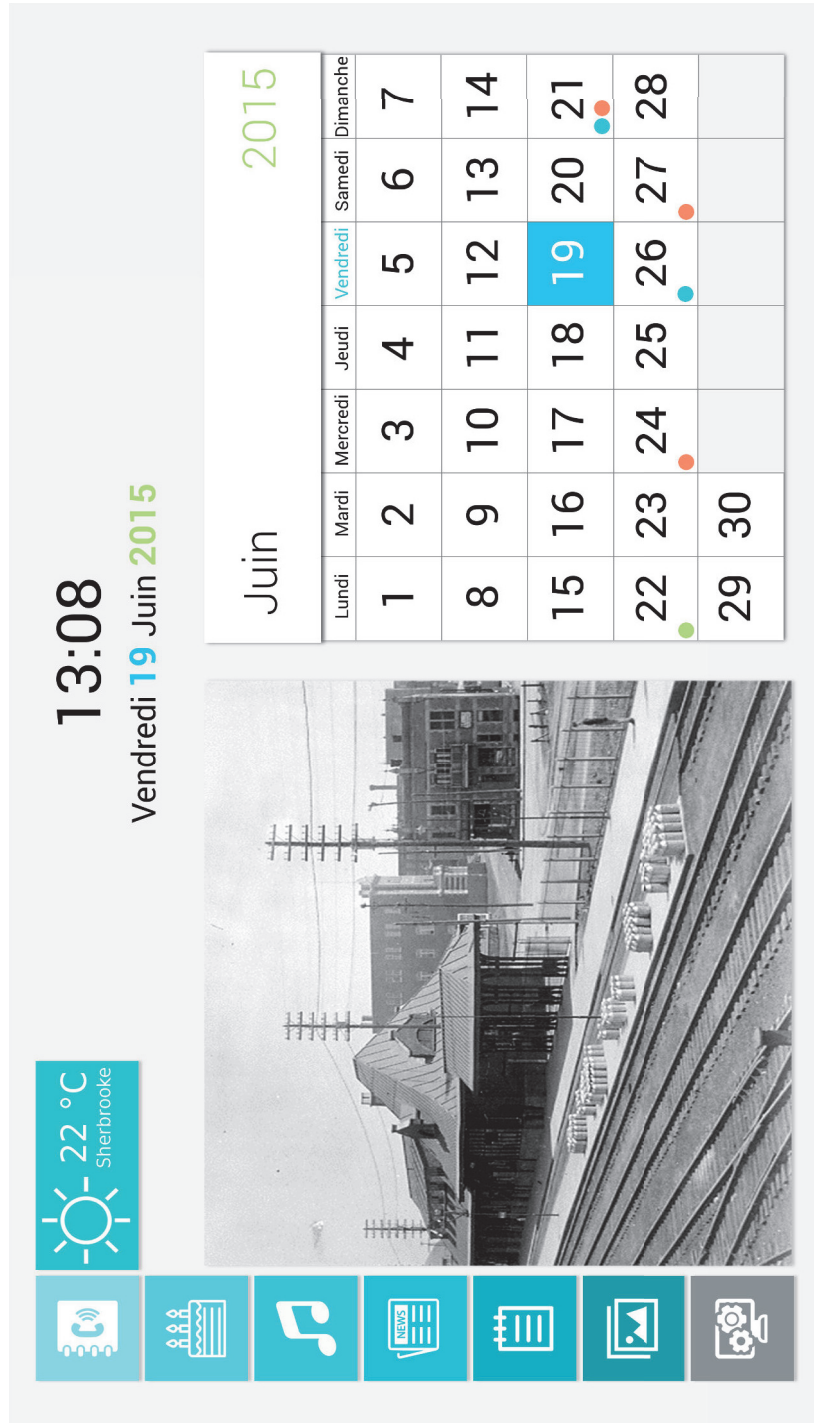


figure A.2 – Maquette du calendrier Amelis au Québec

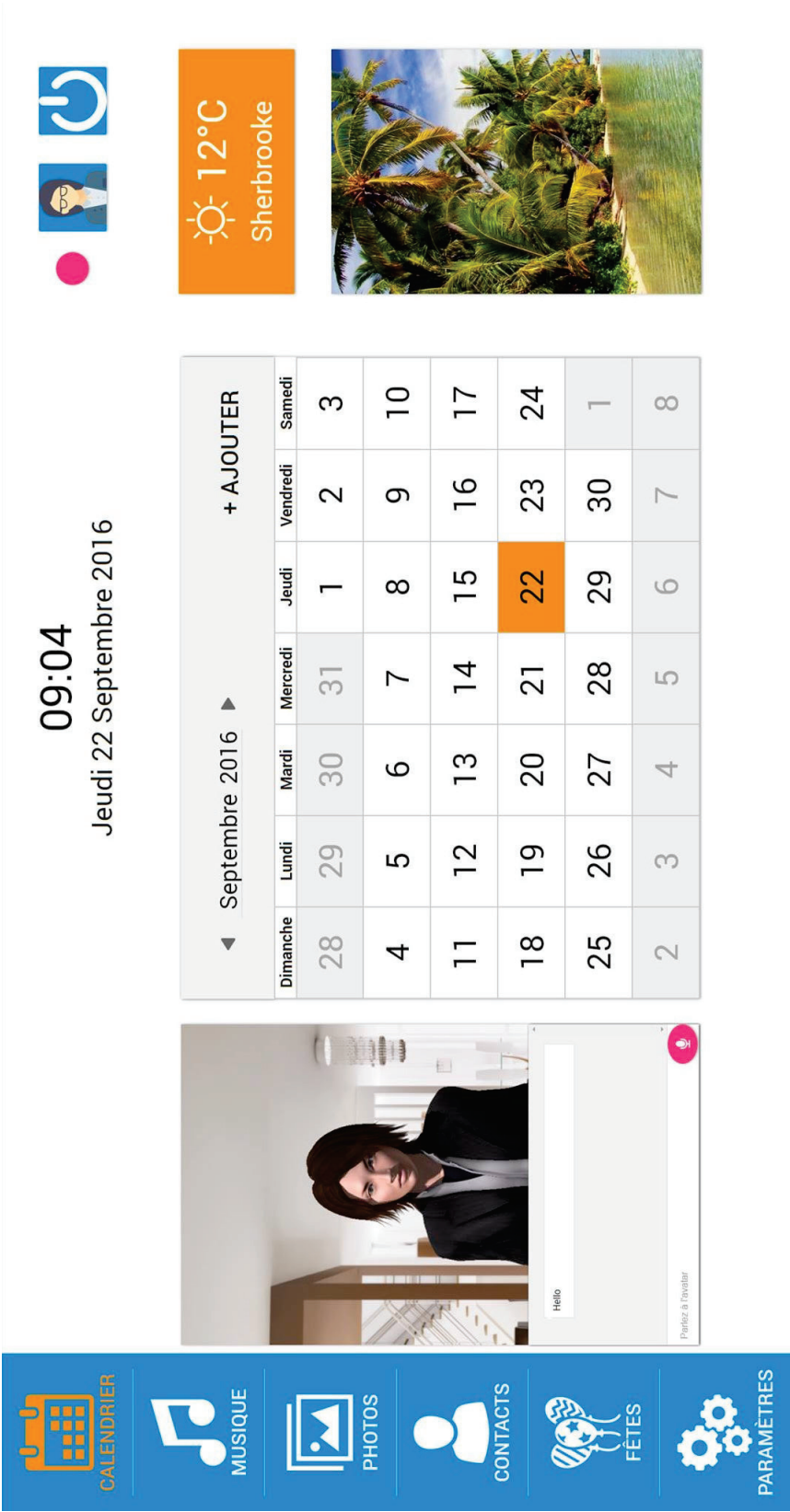


figure A.3 – Maquette commune du calendrier Amelis

Bibliographie

- [1] Thomas KLEINBERGER, Martin BECKER, Eric RAS, Andreas HOLZINGER, and Paul MÜLLER. “Ambient intelligence in assisted living : enable elderly people to handle future interfaces”. In: *International conference on universal access in human-computer interaction*. Springer. 2007, pp. 103–112.
- [2] J.P. HARDY. “Réforme de la tarification des EHPAD (établissements d’hébergement pour personnes âgées dépendantes)”. In: *DGCS – Service des politiques d’appui – Sous-direction des affaires financières et de la modernisation – Bureau de la Gouvernance du secteur* (2017).
- [3] The American Elder Care Research ORGANIZATION. *Technologies to Reduce Care Costs and Allow Safe Aging at Home*. URL: <https://www.payingforseniorcare.com/longtermcare/aging-in-place-technology.html> (visited on 02/03/2017).
- [4] Kurniawan S.H. Fitzpatrick G. et Vetere F. HARLEY D.A. “Age matters : bridging the generation gap through technology-mediated interaction”. In: *Extended abstracts on human factors in computing systems*. Association for Computing Machinery. 2009, pp. 4799–4802.
- [5] A. MCCREADIE C. et Tinker. “The acceptability of assistive technology to older people”. In: *Ageing & Society* (2005), pp. 91–110.
- [6] Kramer M. Pitsch K. et Kopp S. YAGHOUBZADEH R. “Virtual agents as daily assistants for elderly or cognitively impaired people”. In: *International Workshop on Intelligent Virtual Agents*. Springer. 2013, pp. 79–91.

BIBLIOGRAPHIE

- [7] Hiyama A. Miura T. Asakawa C. Hirose M. et Ifukube T. KOBAYASHI M. “Elderly user evaluation of mobile touchscreen interactions”. In: *Conference on Human-Computer Interaction*. Springer. 2011, pp. 83–99.
- [8] S. KELLEY. *The Pros and Cons of Digital Personal Assistants*. URL: <http://www.visionaware.org/info/everyday-living/helpful-products/using-a-computer/comparing-digital-personal-assistants/1245>.
- [9] Loued W.B. Lieb L. Pigot H. Lorrain D. Guérin C. et Chauvin C. PORCHER A. “Projet de conception participative du calendrier Amelis : Placer les aînés au cœur de la conception technologique”. In: *Congrès Age 3.0* (2016).
- [10] V. ORIOL V. et Le Quoc. “A Neural Conversational Model”. In: *Computing Research Repository* (). URL: <http://arxiv.org/abs/1506.05869> (visited on 02/05/2017).
- [11] T.A. SALTHOUSE. “What and when of cognitive aging”. In: *Current directions in psychological science* (2004), pp. 140–144.
- [12] C. BENEFIELD L. E. et Beck. “Reducing the distance in distance-caregiving by technology innovation”. In: *Clinical interventions in aging* (2007), p. 267.
- [13] D. GROVER. *Bots won’t replace apps. Better apps will replace apps*. URL: <http://dangrover.com/blog/2016/04/20/bots-wont-replace-apps.html> (visited on 01/15/2017).
- [14] S. AYALA. “Les personnes âgées face au défi d’utilisation des nouvelles technologies : Étude de l’utilisabilité des interfaces de téléphones portables”. MA thesis. Université de Montréal, 2010.
- [15] Tu Q. et Wang K. SHU Q. “The impact of computer self-efficacy and technology dependence on computer-related technostress : A social cognitive theory perspective”. In: *International Journal of Human-Computer Interaction* (2011), pp. 923–939.
- [16] W. SAMIR K.C. et Lutz. “The human core of the shared socioeconomic pathways : Population scenarios by age, sex and level of education for all countries to 2100”. In: *Global Environmental Change* (2017), pp. 181–192.

BIBLIOGRAPHIE

- [17] Porcher A. Loued W. B. Lieb L. Lorrain D. Guerin C. et Chauvin C. PIGOT H. “ACG2016 : Favoriser l’innovation en recherche sur le vieillissement”. In: *45e Réunion scientifique et éducative annuelle Montréal* (2016).
- [18] M. BEAUDOUIN-LAFON. “Designing Interaction, Not Interfaces”. In: *Proceedings of the Working Conference on Advanced Visual Interfaces*. Association for Computing Machinery. 2004, pp. 15–22.
- [19] Hafez M. Alexandre J.M. et Kheddar A. BENALI-KHOUDJA M. “Tactile interfaces : a state-of-the-art survey”. In: *Int. Symposium on Robotics*. Laboratoire Systèmes Complexes - University of Evry, France. 2004, pp. 23–26.
- [20] Perrottet D. Ladetto Q. et Merminod B. ROUILLER J. “Faciliter le déplacement des aveugles avec une carte numérique et une interface vocale”. In: *Mensuration photogrammetrie génie rural* (2002), pp. 517–522.
- [21] F. STEFFAN I.T. et Tosi. *Ergonomics and design for all*. URL: https://www.researchgate.net/publication/279803487_Ergonomics_and_design_for_all (visited on 02/04/2017).
- [22] Nakano H. et Nagai T. WANNOUS M. “Google Calendar for managing and monitoring the utilization of a web-based laboratory’s resources”. In: *2011 IEEE Global Engineering Education Conference (EDUCON)*. Institute of Electrical and Electronics Engineers. 2011, pp. 210–213.
- [23] D. PRUD’HOMME. “Pour favoriser l’utilisation des technologies par les aînés”. In: *Vie et vieillissement* (2012), pp. 50–52.
- [24] H. DESCHENEUX C. et Pigot. “Interactive calendar to help maintain social interactions for elderly people and people with mild cognitive impairments”. In: *International Conference on Smart Homes and Health Telematics*. Springer. 2009, pp. 117–124.
- [25] N. WEBB. “Rule-based dialogue management systems”. In: *Natural Language Processing Group University of Sheffield*. 2000, pp. 164–169.

BIBLIOGRAPHIE

- [26] Kishore R. Sharman R. et Ramesh R. ZHANG H. “Agile Integration Modeling Language (AIML) : A conceptual modeling grammar for agile integrative business information systems”. In: *Decision Support Systems* (2007), pp. 266–284.
- [27] Yaghoubzadeh R. Kopp S. et Pitsch K. KRAMER M. “A conversational virtual human as autonomous assistant for elderly and cognitively impaired users? Social acceptability and design considerations”. In: *Informatik angepasst an Mensch* (2013).
- [28] Loued W. B. Porcher A. et Pigot H. LIEB L. “Attente des aînés lors d’interactions avec un agent virtuel”. In: *Centre d’Expertise National des Technologies de l’Information et de la Communication , Angers, France* (2017).
- [29] Chetouani M. Tapus A. Bidaud P. et Dupourqué V. GRANATA C. “Voice and graphical-based interfaces for interaction with a robot dedicated to elderly and people with cognitive disorders”. In: *19th International Symposium in Robot and Human Interactive Communication*. Institute of Electrical and Electronics Engineers. 2010, pp. 785–790.
- [30] Möller S. Engelbrecht K. Kühnel C. Schleicher R. Naumann A. et Wolters M. GÖDDE F. “Study of a speech-based smart home system with older users”. In: *Conference : International workshop on intelligent user interfaces for ambient assisted living*. 2008, pp. 17–22.
- [31] Cohen M. et Giangola J.P. BALOGH J. “Minimizing cognitive load”. In: *Voice User Interface Design* (2004).
- [32] K. ADAMS. “Speak and Write : Assessing Spoken Production in Large Classes”. In: *National Diet Library* (2014).
- [33] A. MERILAMPI S. et Sirkka. *Introduction to Smart EHealth and ECare Technologies*.
- [34] D. ; Laurent R. ; Piotr B. ROMAIN L. ; Jan. “Cooking Coach Spoken/Multimodal Dialogue Systems”. In: *Conference : Proceedings of the International Joint Conference on Artificial Intelligence Workshop on Cooking with Computers* (2013).

BIBLIOGRAPHIE

- [35] C. LAMBERTSSON. *Expectations of Privacy in Voice Interaction—A Look at Voice Controlled Bank Transactions*. Royal Institute of Technology, Sweden, 2017.
- [36] Young V. Boger J. et Mihailidis A. HAMILL M. “Development of an automated speech recognition interface for personal emergency response systems”. In: *Journal of NeuroEngineering and Rehabilitation* (2009).
- [37] D. LUZZATI. “Pour une typologie des tâches en dialogue homme-machine”. In: *Psychologie du dialogue homme-machine en langage naturel* (1996). Ed. by Europia PRODUCTIONS, pp. 107–115.
- [38] Bangor A.W. et Bushey R.R. JOSEPH K.M. *Call routing from manual to automated dialog of interactive voice response system*. American Telephone Telegraph Laboratories, 2004.
- [39] L.R. JUANG B.H. et Rabiner. “Automatic speech recognition : a brief history of the technology development”. In: *Georgia Institute of Technology*. (2005), p. 67.
- [40] A. DROPO J. et Acero. “Noise robust speech recognition with a switching linear dynamic model”. In: *International Conference on Acoustics, Speech and Signal Processing*. 2004.
- [41] C. VERVERIDIS D. et Kotropoulos. “Emotional speech recognition : Resources, features, and methods”. In: *Speech communication* (2006), pp. 1162–1181.
- [42] Droppo J. Huang X. Seide F. Seltzer M. Stolcke A. Yu D. et XIONG W. and G. ZWEIG. “Achieving Human Parity in Conversational Speech Recognition”. In: *Computing Research Repository* (). URL: <http://arxiv.org/abs/1610.05256> (visited on 01/21/2017).
- [43] O. LEMON. “Conversational interfaces”. In: 2012, pp. 1–4.
- [44] H. JIANG. “Confidence measures for speech recognition : A survey”. In: *Speech communication* (2005), pp. 455–470.
- [45] F. JESCHKE B. et Reh. *Speech dialogue system for dialogue interruption and continuation control*. Harman Becker Automotive Systems, 2008.

BIBLIOGRAPHIE

- [46] Busayapongchai S. Giass J. Goddeau D. Hethetingron L. Hurley E. Pao C. Polifroni J. Seneff S. et Zue V. MENG He. “WHEELS : A conversational system in the automobile classifieds domain”. In: Institute of Electrical and Electronics Engineers. 1996, pp. 542–545.
- [47] T. LEMEUNIER. “L’intentionnalité communicative dans le dialogue homme-machine en langue naturelle”. PhD thesis. 2000, p. 235.
- [48] E.S. SHAWAR A. et Atwell. “A chatbot system as a tool to animate a corpus”. In: *International Computer Archive of Modern and Medieval English Journal* (2005), pp. 5–24.
- [49] C. PEARL. *Designing Voice User Interfaces : Principles of Conversational Experiences*. 2016.
- [50] E. CONNOLLY. *Principles of bot design*. URL: <https://blog.intercom.com/principles-bot-design/> (visited on 01/15/2017).
- [51] Jaramillo D. McCobb G. et Wilson L.R. ATIVANICHAYAPHONG S. *Method of enhancing voice interactions using visual messages*. International Business Machines Corp, 2011.
- [52] F.F. GALANXHI H. et Nah. “Deception in cyberspace : A comparison of text-only vs. avatar-supported medium”. In: *International journal of human-computer studies* (2007), pp. 770–783.
- [53] von der Pütten A. Krämer N. et Gratch J. APPEL J. “Does humanity matter ? Analyzing the importance of social cues and perceived agency of a computer system for the emergence of social reactions during human-computer interaction”. In: *Advances in Human-Computer Interaction* (2012), p. 13.
- [54] I. QIU L. et Benbasat. “An investigation into the effects of Text-To-Speech voice and 3D avatars on the perception of presence and flow of live help in electronic commerce”. In: *Transactions on Computer-Human Interaction (TOCHI)* (2005), pp. 329–355.
- [55] Míkovec Z. et Slavík P. KUNC L. “Avatar and dialog turn-yielding phenomena”. In: *International Journal of Technology and Human Interaction (IJTHI)* (2013), pp. 66–88.

BIBLIOGRAPHIE

- [56] A. XENON. *Écrire un dialogue*. URL: <http://ffnetmodedemploi.free.fr/dialogue.php> (visited on 01/16/2017).
- [57] S. MINKER W. et Bennacef. *Speech and human-machine dialog*. Vol. 770. 2004.
- [58] Y.H. af SEGERSTAD. *Use and adaptation of written language to the conditions of computer-mediated communication*. 2002.
- [59] J.M. CARROLL. “Human-computer interaction : psychology as a science of design”. In: *Annual review of psychology* (1997), pp. 61–83.
- [60] S. NASS C.I. et Brave. *Wired for speech : How voice activates and advances the human-computer relationship*. 2005.
- [61] Breazeal C. et Norton M.I. SIEGEL M. “Persuasive robotics : The influence of robot gender on human behavior”. In: Institute of Electrical and Electronics Engineers. 2009, pp. 2563–2568.
- [62] P. NOLLER. “Misunderstandings in marital communication : A study of couples’ nonverbal communication.” In: *Journal of Personality and Social Psychology* (1980), p. 1135.
- [63] Frossard P. Kokiopoulou E. et Verscheure O. AMINI L. *Method and system for robust pattern matching in continuous speech for spotting a keyword of interest using orthogonal matching pursuit*. International Business Machines Corp, 2016.
- [64] C. PEARL. *Designing voice user interfaces*. 2016.
- [65] E.A. BRENNAN S.E. et Hulteen. “Interaction and feedback in a spoken language system : A theoretical framework”. In: *Knowledge-based systems* (1995), pp. 143–151.
- [66] D. SCHNELLE-WALKA. “A pattern language for error management in voice user interfaces”. In: Association for Computing Machinery. 2010, p. 8.
- [67] B. WEBER I. et Tausend. “A three-tiered confidence model for revising logical theories”. In: Citeseer. 1994, pp. 391–402.
- [68] Karpicke J.A. Keen D.M. et Lively B.L. COHRS P.W. *Method of implicit confirmation for automatic speech recognition*. Touch Controls Inc, 1999.

BIBLIOGRAPHIE

- [69] Hanna P. Liu X. et McTear M. O'NEILL I. "The queen's communicator : An object-oriented dialogue manager". In: San Franscic, California. 2003.
- [70] J. FRULLA R.E. et Kirby. *Voice input system and method of using same*. Touch Controls Inc, 2002.
- [71] A. BARKHUUS L. et Dey. "Is context-aware computing taking control away from the user? Three levels of interactivity examined". In: Springer. 2003, pp. 149–156.
- [72] J.K. BAKER. "Trainable grammars for speech recognition". In: *The Journal of the Acoustical Society of America* (1979), S132–S132.
- [73] D. PEREZ-MARIN. *Conversational Agents and Natural Language Interaction : Techniques and Effective Practices : Techniques and Effective Practices*. 2011.
- [74] S. HAYKIN. "A comprehensive foundation". In: *Neural Networks* (2004), p. 41.
- [75] Vinyals O. et Le Q.V. SUTSKEVER I. "Sequence to sequence learning with neural networks". In: 2014, pp. 3104–3112.
- [76] J.H. DIANE. "Latent Dirichlet Allocation for Text, Images, and Music". In: *Department of Computer Science University of California, San Diego* (2009), p. 1.
- [77] D. MCNEAL M.L. et Newyear. "Chatbot creation options". In: *Library Technology Reports* (2013), pp. 11–17.
- [78] "the newsletter of the society for the study of artificial intelligence and simulation of behaviour". In: ed. by Artificial INTELLIGENCE and Simulation of BEHAVIOUR. 2016.
- [79] T. HALLER Em. et Rebedea. "Designing a chat-bot that simulates an historical figure". In: Institute of Electrical and Electronics Engineers. 2013, pp. 582–589.
- [80] Siddharth GUPTA, Deep BORKAR, Chevelyn DE MELLO, and Saurabh PATIL. "An E-Commerce Website based Chatbot". In: *Citeseer* (2015).
- [81] D. HIEMSTRA. "Information retrieval models". In: *Information Retrieval : searching in the 21st Century* (2009), pp. 1–17.

BIBLIOGRAPHIE

- [82] R. SALAKHUTDINOV. “Learning deep generative models”. In: *Annual Review of Statistics and Its Application* (2015), pp. 361–385.
- [83] Desouza P.V. Mercer R.L. Pietra V. et Lai J.C. BROWN P.F. “Class-based n-gram models of natural language”. In: *Computational linguistics* (1992), pp. 467–479.
- [84] A. STOLCKE. “SRILM-an extensible language modeling toolkit.” In: Stanford Research Institute. 2002.
- [85] D. BRITZ. *Deep learning for chatbots, part 1 introduction*. URL: <http://www.wildml.com/2016/04/deep-learning-for-chatbots-part-1-introduction/> (visited on 01/19/2017).
- [86] et Pigot H. LOUED W. B. “Emotional Virtual Agent to Improve Ageing in Place with Technology”. In: *6th International Conference on Digital Health Conference* (2016).
- [87] Newell A.F. et Zajicek M. GREGOR P. “Designing for dynamic diversity : interfaces for older people”. In: *Proceedings of the fifth international ACM conference on Assistive technologies*. Association for Computing Machinery. 2002, pp. 151–156.
- [88] Porcher A. Lieb L. Pigot H. Lorrain D. Chauvin C. et Guerin C. LOUED W. B. “Amelis : concevoir un agent virtuel expressif intégré à un calendrier interactif pour des aînés”. In: *Workshop sur les “Affects, Compagnons Artificiels et Interactions”, Brest, France* (2016).
- [89] C. MANFREDI G. et Gribaudo. *Virtual Assistant With Real-Time Emotions*. Kallideas, 2008.
- [90] Roukos S. Ward T. et Zhu W.J. PAPINENI K. “BLEU : a method for automatic evaluation of machine translation”. In: *Proceedings of the 40th annual meeting on association for computational linguistics*. Association for Computational Linguistics. 2002, pp. 311–318.
- [91] A. LAVIE. *Evaluating the output of machine translation systems*. URL: https://www.researchgate.net/publication/266096165_Evaluating_the_Output_of_Machine_Translation_Systems (visited on 01/24/2017).

BIBLIOGRAPHIE

- [92] McLean D. Bandar Z.A. LI Y. and K. O'SHEA J.D. et Crockett. "Sentence similarity based on semantic nets and corpus statistics". In: *Transactions on knowledge and data engineering* (2006), pp. 1138–1150.
- [93] Phillips N.A. Collin I. Cummings J.L. et Chertkow H. NASREDDINE Z.S. "The Montreal Cognitive Assessment, MoCA : a brief screening tool for mild cognitive impairment". In: *Journal of the American Geriatrics Society* (2005), pp. 695–699.
- [94] J. ABDUL-KADER S. et Woods. "Survey on chatbot design techniques in speech conversation systems". In: *International Journal of Advanced Computer Science and Applications* (2015), pp. 72–80.
- [95] Goldie A. Luong T et Le Quoc V. BRITZ D. "Massive Exploration of Neural Machine Translation Architectures". In: *Computing Research Repository* (). URL: <http://arxiv.org/abs/1703.03906> (visited on 04/18/2017).
- [96] Kim Y. Deng Y. Senellart J. et Rush A.M. KLEIN G. "OpenNMT : Open-Source Toolkit for Neural Machine Translation". In: *Computing Research Repository* (2017-04-18). URL: <http://arxiv.org/abs/1701.02810>.
- [97] Schuster M. Chen Z. Le Quoc V. Norouzi M. Macherey W. Krikun M. Cao Y. Gao Q. et Macherey K. WU Y. "Google's Neural Machine Translation System : Bridging the Gap between Human and Machine Translation". In: *Computing Research Repository* (). URL: <http://arxiv.org/abs/1609.08144> (visited on 01/29/2017).
- [98] Y. ODASHI. *Collection of Evaluation Metrics and Algorithms for Machine Translation*. URL: <https://github.com/odashi/mteval> (visited on 04/05/2017).
- [99] P. SUJIT. *Worked examples from the NLTK Book*. URL: <https://github.com/sujitpal/nltk-examples/> (visited on 04/07/2017).
- [100] E. GAMMA. *Design patterns : elements of reusable object-oriented software*. Ed. by ADDISON-WESLEY. 1995.

BIBLIOGRAPHIE

- [101] Zaremba W. Sutskever I. Bruna J. Erhan D. Goodfellow I. et Fergus R. SZEGEDY C. “Intriguing properties of neural networks”. In: *Computing Research Repository* (). URL: <http://arxiv.org/abs/1312.6199> (visited on 02/08/2017).
- [102] Nandan N. et Sun A. HAN J. “Did You Know? A Rule-Based Approach to Finding Similar Questions on Online Health Forums”. In: Institute of Electrical and Electronics Engineers. 2015, pp. 513–514.
- [103] Alleva F. Hon H. Hwang M. Lee K. et Rosenfeld R. HUANG X. “The SPHINX-II speech recognition system : an overview”. In: *Computer Speech & Language* (1993), pp. 137–148.